



การทำเหมืองข้อมูลสำหรับการค้นหาแนวโน้มราคาหลักทรัพย์
ด้วยกรอบความน่าจะเป็น

โดย

นายณวัฏมิ พานิชย์วิไล

โครงการพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์บัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2565

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

การทำเหมืองข้อมูลสำหรับการค้นหาแนวโน้มราคาหลักทรัพย์

ด้วยกรอบความน่าจะเป็น

โดย

นายณฐภูมิ พานิชย์วิไล

โครงการพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์บัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2565

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

MINING STOCK PRICE TRENDS USING PROBABILISTIC FRAMEWORK

BY

MISTER NAWAPOOM PANICHWILAI

A FINAL-YEAR PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE

COMPUTER SCIENCE

FACULTY OF SCIENCE AND TECHNOLOGY

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2022

COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

รายงานโครงการพิเศษ

ของ

นายณฐกร พานิชย์วิไล

เรื่อง

การทำเหมืองข้อมูลสำหรับการค้นหาแนวโน้มราคาหลักทรัพย์ด้วยกรอบความน่าจะเป็น

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

เมื่อ วันที่ วันที่ ชื่อเต็มของเดือน พ.ศ. 2565

อาจารย์ที่ปรึกษา

(อ.ดร. ฐานา บุญชู)

อาจารย์ที่ปรึกษาร่วม

(อ.ดร. ศาตนาฏ กิจศิริานุวัตร)

กรรมการสอบโครงการพิเศษ

(ผศ.ดร. เสาวลักษณ์ วรรณภา)

กรรมการสอบโครงการพิเศษ

(ผศ.ดร. วิลาวรรณ รักผกาวงศ์)

กรรมการสอบโครงการพิเศษ

(รศ.ดร. ณัฐชนน หงส์วริทธิ์ธร)

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

รายงานโครงการพิเศษ

ของ

นายณัฐภูมิ พานิชย์วิไล

เรื่อง

การทำเหมืองข้อมูลสำหรับการค้นหาแนวโน้มราคาหลักทรัพย์ด้วยกรอบความน่าจะเป็น

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

เมื่อ วันที่ วันที่ ชื่อเต็มของเดือน พ.ศ. 2565

อาจารย์ที่ปรึกษา

(อ.ดร. ฐาปนา บุญชู)

อาจารย์ที่ปรึกษาร่วม

(อ.ดร. ศาตนาฏ กิจศิริานุวัตร)

กรรมการสอบโครงการพิเศษ

(ผศ.ดร. เสาวลักษณ์ วรรณภา)

กรรมการสอบโครงการพิเศษ

(ผศ.ดร. วิลาวรรณ รักผกาวงศ์)

กรรมการสอบโครงการพิเศษ

(รศ.ดร. ณัฐชนน หงส์วริทธิ์ธร)

หัวข้อโครงการพิเศษ	การทำเหมืองข้อมูลสำหรับการค้นหาแนวโน้มราคา
	หลักทรัพย์ด้วยกรอบความน่าจะเป็น
ชื่อผู้เขียน	นายณฐภูมิ พานิชย์วิไล
ชื่อปริญญา	วิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาวิทยาการคอมพิวเตอร์
	คณะวิทยาศาสตร์และเทคโนโลยี
	มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาโครงการพิเศษ	อ.ดร. ฐาปนา บุญชู
อาจารย์ที่ปรึกษาโครงการพิเศษร่วม	อ.ดร. ศาตนาฏ กิจศิริานุวัตร
ปีการศึกษา	2565

บทคัดย่อ

โครงการพิเศษนี้มีเป้าหมายในการศึกษาและพัฒนากรอบวิธีการทำนายแนวโน้มราคาหลักทรัพย์แบบรายวัน (Daily trends) ผ่านเทคนิคการทำเหมืองข้อมูลโดยใช้วิธีค้นพบโมทีฟในอนุกรมเวลา (Data mining) เพื่อค้นหารูปแบบของราคาในอดีตที่ซ่อนอยู่มาใช้เป็นตัวแปรในการวิเคราะห์และสร้างวิธีการทำนายผล โดยทำการศึกษาและเปรียบเทียบกับเทคนิควิธีการพื้นฐานที่ใช้ตัวชี้วัดทางเทคนิคเป็นตัวแปรในงานที่เกี่ยวข้องส่วนใหญ่ สำหรับวิธีการทำนายผลจะทำการศึกษาโดยใช้กรอบความน่าจะเป็นมาสร้างวิธีการทำนายผ่านวิธีการวิเคราะห์ทางสถิติ (Statistical analysis) และวิธีการเรียนรู้ของเครื่อง ได้แก่ วิธีจำแนกแบบเบย์อย่างง่าย (Naïve bayes classifier) และวิธีจำแนกต้นไม้การตัดสินใจ (Decision tree classifier) เพื่อทำนายแนวโน้มการเปลี่ยนแปลงราคาหลักทรัพย์แบบรายวัน โดยมีเป้าหมายเพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพความแม่นยำของการทำนายผลด้วยเทคนิคที่แตกต่างกัน

จากผลการทดลองพบว่า วิธีการทำนายแนวโน้มราคาหลักทรัพย์ผ่านเทคนิควิธีการทำเหมืองข้อมูลมีค่า Mean Accuracy สูงกว่าเทคนิควิธีการพื้นฐานโดยเฉลี่ยอยู่ที่ 1-3% และค่า Mean Precision สูงกว่าเทคนิควิธีการพื้นฐานโดยเฉลี่ยอยู่ที่ 1-8% โดยสำหรับเทคนิควิธีการทำเหมืองข้อมูล วิธีการจำแนกแบบเบย์อย่างง่าย (Naïve bayes classifier) เป็นวิธีที่ให้ค่า Accuracy และ Precision เฉลี่ยสูงที่สุด คือ 0.6022 และ 0.6347 ตามลำดับ และในส่วนของเทคนิควิธีการพื้นฐาน

วิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multilayer perceptron classifier) เป็นวิธีที่ให้ค่า Accuracy เฉลี่ยสูงสุด คือ 0.5725 และวิธีจำแนกต้นไม้การตัดสินใจ (Decision tree classifier) เป็นวิธีที่ให้ค่า Precision เฉลี่ยสูงสุด คือ 0.5995 ซึ่งเมื่อพิจารณาโดยรวมจะเห็นได้ว่าวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูลที่ได้ทดลองศึกษาสามารถแสดงค่าความถูกต้องและค่าความแม่นยำได้สูงขึ้นเมื่อเทียบกับวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค

คำสำคัญ: แนวโน้ม, เทคนิควิธีการทำเหมืองข้อมูล, รูปแบบของราคา, เทคนิควิธีการพื้นฐาน, ตัวชี้วัดทางเทคนิค, ตัวแปร, กรอบความน่าจะเป็น, วิธีการเรียนรู้ของเครื่อง, ข้อมูลอนุกรมเวลา

Thesis Title	Mining Stock Price Trends using Probabilistic Framework
Author	MR. Nawapoom Panichwilai
Degree	Bachelor of Science
Major Field/Faculty/University	Computer Science Faculty of Science and Technology Thammasat University
Project Advisor	Dr. Thapana Boonchoo
Project Co-Advisor	Dr. Satanat Kitsiranuwat
Academic Years	2022

ABSTRACT

This special project aims to study and develop a framework for predicting daily stock price trends using data mining techniques with time-series motif discovery method to find hidden similar price patterns in the past to be used as variables for analyzing and generating predictive modeling. The study was conducted and compared with baseline techniques using technical indicators as variables in most relevant studies. As for predictive modeling, the study uses a probabilistic framework to build a model through statistical analysis and machine learning models such as Naïve bayes classifier and Decision tree classifier for predicting daily stock price trends. The goal was to study and compare the performance of different techniques.

From the experimental results, it was show that using data mining techniques has a mean accuracy about 1-3% higher than baseline techniques and mean precision also higher than baseline technique about 1-8%. For data mining techniques, Naïve bayes classifier show the highest mean accuracy and precision about 0.6022 and 0.6347 respectively. For baseline techniques, Multilayer perceptron classifier show the highest mean accuracy about 0.5725 and Decision tree classifier show the highest mean

precision about 0.5995. In overall, it can be seen that predicting daily stock price trends using data mining technique show a better performance with higher of accuracy and precision compared to baseline techniques using technical indicators.

Keywords: Trends, Data mining techniques, Price patterns, Baseline techniques, Technical Indicators, Features, Probabilistic framework, Machine learning, Time-series data

กิตติกรรมประกาศ

โครงการฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดีอันเนื่องมาจากความกรุณาช่วยเหลือจากอาจารย์ที่ปรึกษาโครงการพิเศษ อ.ดร.ฐาปนา บุญชู และอาจารย์ที่ปรึกษาโครงการพิเศษร่วม อ.ดร. ศาตนาฏ กิจศิริานุวัตร ที่ได้กรุณาสละเวลามาช่วยให้คำปรึกษาและแนะนำในด้านแนวคิดของโครงการ เนื้อหาของโครงการ ตลอดจนช่วยแก้ไขข้อบกพร่องต่างๆในโครงการฉบับนี้จนเสร็จสิ้นและครบถ้วนสมบูรณ์

สุดท้ายผู้เขียนขอกล่าวขอบคุณทุกท่านที่ได้กล่าวถึงและไม่ได้กล่าวถึงในที่นี้ที่ได้ให้ความช่วยเหลือและเป็นกำลังใจให้ผู้เขียนในการจัดทำโครงการฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

นายนวนภูมิ พานิชย์วิไล

สารบัญ

หน้า

บทคัดย่อ.....	1
ABSTRACT.....	3
กิตติกรรมประกาศ	5
สารบัญ	6
สารบัญตาราง	9
สารบัญภาพ	11
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของโครงการ.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตของโครงการ	3
1.4 ประโยชน์ของโครงการ	4
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	5
2.1 แนวคิดทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 Domain Knowledge.....	5
2.1.2 Computer Science Knowledge	8
2.2 งานวิจัยที่เกี่ยวข้อง.....	22
2.3 เทคโนโลยีที่เกี่ยวข้อง	26

2.3.1 ภาษาโปรแกรมไพทอน (Python Programming Language).....	26
2.3.2 ไลบรารีที่เกี่ยวข้อง (Library)	26
2.3.3 เครื่องมือ (Tools).....	28
บทที่ 3 วิธีการวิจัย.....	29
3.1 ภาพรวมของการศึกษา	29
3.2 วิธีการศึกษา	30
3.2.1 ขั้นตอนการรวบรวมข้อมูล (Data Collection).....	30
3.2.2 ขั้นตอนการเตรียมข้อมูล (Data Preprocessing)	31
3.2.3 ขั้นตอนการทำเหมืองข้อมูล (Data Mining).....	32
3.2.4 ขั้นตอนการสร้างวิธีการทำนายผล (Predictive Modeling)	34
3.2.5 ขั้นตอนการวัดประสิทธิภาพ (Model Evaluation)	38
3.2.6 ขั้นตอนการปรับค่าพารามิเตอร์ (Parameter tuning).....	39
3.2.7 ขั้นตอนการวิเคราะห์แนวโน้มหลักทรัพย์รายวันผ่านแพลตฟอร์มซื้อขายจำลอง	39
บทที่ 4 ผลการดำเนินงาน.....	40
4.1 การเก็บข้อมูล.....	40
4.2 ผลการทดลองของเทคนิควิธีการทำเหมืองข้อมูล (Data mining techniques).....	40
4.2.1 การค้นพบโมทีฟในอนุกรมเวลา (Time series motif discovery).....	40
4.2.2 หลักทรัพย์ KBANK.BK (กลุ่มธุรกิจการเงิน).....	41
4.2.3 หลักทรัพย์ PTTEP.BK (กลุ่มทรัพยากร).....	45
4.2.4 หลักทรัพย์ CPALL.BK (กลุ่มบริการ).....	49
4.2.5 หลักทรัพย์ DELTA.BK (กลุ่มเทคโนโลยี).....	53
4.2.6 หลักทรัพย์ SCC.BK (กลุ่มอสังหาริมทรัพย์และการก่อสร้าง)	57

4.3	ผลการทดลองของวิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค	61
4.3.1	การคำนวณตัวชี้วัดทางเทคนิค (Technical Indicators)	61
4.3.2	การกำหนดค่าไฮเปอร์พารามิเตอร์และเทคนิคการดึงคุณลักษณะของโมเดล (Hyperparameter tuning and Feature extraction techniques)	62
4.3.1	การคัดเลือกฟีเจอร์ (Feature selection)	63
4.3.2	หลักทรัพย์ KBANK.BK (กลุ่มธุรกิจการเงิน).....	66
4.3.3	หลักทรัพย์ PTTEP.BK (กลุ่มทรัพยากร).....	69
4.3.4	หลักทรัพย์ CPALL.BK (กลุ่มบริการ).....	72
4.3.5	หลักทรัพย์ DELTA.BK (กลุ่มเทคโนโลยี).....	75
4.3.6	หลักทรัพย์ SCC.BK (กลุ่มอสังหาริมทรัพย์และการก่อสร้าง)	78
4.4	การศึกษาประสิทธิภาพและเปรียบเทียบ (Model Evaluation and Comparison)	81
4.4.1	ประสิทธิภาพวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูล	81
4.4.2	ประสิทธิภาพวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐาน.....	82
4.5	สรุปผลการเปรียบเทียบ.....	83
4.6	ผลการวิเคราะห์การทำนายแนวโน้มแบบรายวันผ่านแพลตฟอร์มซื้อขายหลักทรัพย์จำลอง	84
บทที่ 5	สรุป.....	86
5.1	สรุปผลการทดลอง	86
5.2	แนวคิดต่อยอดและการปรับปรุงในอนาคต	88
	แหล่งอ้างอิง	89

สารบัญตาราง

หน้า

ตารางที่ 4.1	รายการข้อมูลตลาดหลักทรัพย์ที่ใช้ในศึกษา	40
ตารางที่ 4.2	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ KBANK.BK.....	44
ตารางที่ 4.3	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ PTTEP.BK.....	48
ตารางที่ 4.4	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ CPALL.BK.....	52
ตารางที่ 4.5	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ DELTA.BK	56
ตารางที่ 4.6	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ SCC.BK.....	60
ตารางที่ 4.7	ประเภทของอินดิเคเตอร์	61
ตารางที่ 4.8	เทคนิควิธีการดึงคุณลักษณะต่างๆบนวิธีจำแนกแบบเบย์อย่างง่าย.....	62
ตารางที่ 4.9	ค่าไฮเปอร์พารามิเตอร์บนโมเดลวิธีจำแนกต้นไม้การตัดสินใจ	62
ตารางที่ 4.10	ค่าไฮเปอร์พารามิเตอร์บนโมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น	62
ตารางที่ 4.11	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ KBANK.BK.....	68
ตารางที่ 4.12	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ PTTEP.BK.....	71
ตารางที่ 4.13	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ CPALL.BK	74
ตารางที่ 4.14	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ DELTA.BK	77
ตารางที่ 4.15	ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ SCC.BK.....	80
ตารางที่ 4.16	ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีการวิเคราะห์ทางสถิติ	81

ตารางที่ 4.17 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกแบบเบย์อย่างง่าย	81
ตารางที่ 4.18 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกต้นไม้การตัดสินใจ	81
ตารางที่ 4.19 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีการวิเคราะห์ทางสถิติ	82
ตารางที่ 4.20 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกแบบเบย์อย่างง่าย	82
ตารางที่ 4.21 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น	82
ตารางที่ 4.22 ผลการวิเคราะห์แนวโน้มราคาหลักทรัพย์ KBANK.BK แบบรายวัน	84
ตารางที่ 4.23 ผลการวิเคราะห์แนวโน้มราคาหลักทรัพย์ DELTA.BK แบบรายวัน	85

สารบัญภาพ

หน้า

ภาพที่ 2.1 ตัวอย่างวิธีการค้นพบโมทีฟในข้อมูลอนุกรมเวลา.....	9
ภาพที่ 2.2 ตัวอย่างวิธีการค้นพบดิสคอร์ดในข้อมูลอนุกรมเวลา	10
ภาพที่ 2.3 เปรียบเทียบการวัดระยะห่างระหว่าง Euclidean Distance กับ Dynamic Time Warping	11
ภาพที่ 2.4 เส้นทางการวอร์ป (warping path).....	12
ภาพที่ 2.5 Pseudocode แสดงขั้นตอนวิธีการทำงานของมาตรวัดแบบ Dynamic Time Warping.....	13
ภาพที่ 2.6 ตัวอย่างแผนภูมิต้นไม้การตัดสินใจ	18
ภาพที่ 3.1 แผนภาพแสดงภาพรวมของการศึกษา	29
ภาพที่ 3.2 ตัวอย่างข้อมูลดิบของรายการหลักทรัพย์ (Raw data)	30
ภาพที่ 3.3 ตัวอย่าง Sliding window technique บนข้อมูลอนุกรมเวลา.....	32
ภาพที่ 3.4 ตัวอย่างวิธีการค้นพบโมทีฟบนข้อมูลอนุกรมเวลา	33
ภาพที่ 3.5 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีการวิเคราะห์ทางสถิติ	34
ภาพที่ 3.6 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีจำแนกแบบเบย์อย่างง่าย	36
ภาพที่ 3.7 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีจำแนกต้นไม้การตัดสินใจ	37
ภาพที่ 3.8 แพลตฟอร์ม Streaming Click2Win	39
ภาพที่ 4.1 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ KBANK.BK.....	41
ภาพที่ 4.2 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ KBANK.BK	42

ภาพที่ 4.3 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ KBANK.BK.....	43
ภาพที่ 4.4 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ KBANK.BK.....	44
ภาพที่ 4.5 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีการ วิเคราะห์ทางสถิติบนหลักทรัพย์ PTTEP.BK.....	45
ภาพที่ 4.6 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ PTTEP.BK.....	46
ภาพที่ 4.7 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ PTTEP.BK.....	47
ภาพที่ 4.8 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ PTTEP.BK.....	48
ภาพที่ 4.9 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีการ วิเคราะห์ทางสถิติบนหลักทรัพย์ CPALL.BK	49
ภาพที่ 4.10 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ CPALL.BK.....	50
ภาพที่ 4.11 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ CPALL.BK	51
ภาพที่ 4.12 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ CPALL.BK	52
ภาพที่ 4.13 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีการ วิเคราะห์ทางสถิติบนหลักทรัพย์ DELTA.BK	53
ภาพที่ 4.14 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ DELTA.BK.....	54
ภาพที่ 4.15 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธี จำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ DELTA.BK	55

ภาพที่ 4.16 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ DELTA.BK.....	56
ภาพที่ 4.17 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ SCC.BK.....	57
ภาพที่ 4.18 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ SCC.BK.....	58
ภาพที่ 4.19 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ SCC.BK.....	59
ภาพที่ 4.20 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ SCC.BK.....	60
ภาพที่ 4.21 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ KBANK.BK.....	63
ภาพที่ 4.22 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ PTTEP.BK.....	64
ภาพที่ 4.23 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ CPALL.BK.....	64
ภาพที่ 4.24 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ DELTA.BK.....	65
ภาพที่ 4.25 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ SCC.BK.....	65
ภาพที่ 4.26 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างควบคู่กับเทคนิควิธีที่แตกต่างกันบนหลักทรัพย์ KBANK.BK.....	66
ภาพที่ 4.27 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ KBANK.BK.....	68
ภาพที่ 4.28 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่แตกต่างกันบนหลักทรัพย์ PTTEP.BK.....	69
ภาพที่ 4.29 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ PTTEP.BK.....	71
ภาพที่ 4.30 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่แตกต่างกันบนหลักทรัพย์ CPALL.BK.....	72
ภาพที่ 4.31 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ CPALL.BK.....	74

ภาพที่ 4.32 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่แตกต่าง กันบนหลักทรัพย์ DELTA.BK.....	75
ภาพที่ 4.33 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ DELTA.BK.....	77
ภาพที่ 4.34 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่แตกต่าง กันบนหลักทรัพย์ SCC.BK.....	78
ภาพที่ 4.35 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ SCC.BK	80

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของโครงการ

การทำนายแนวโน้มราคาหลักทรัพย์ (Stock price trends forecasting) เป็นหนึ่งในหัวข้อที่ได้รับความสนใจและมีการศึกษากันมาหลายทศวรรษจากนักวิจัยในหลากหลายสาขา เนื่องจากข้อมูลราคาหลักทรัพย์นั้นเป็นข้อมูลอนุกรมเวลาที่มีความซับซ้อน มีความผันผวนค่อนข้างสูง และมีลักษณะไม่เป็นเชิงเส้น ซึ่งก่อให้เกิดความท้าทายอย่างยิ่งในการวิเคราะห์และทำนายแนวโน้มของราคาที่จะเกิดขึ้นในอนาคต นักวิจัยและผู้ลงทุนจำนวนมากจึงได้พยายามค้นหาเทคนิควิธีต่างๆ ในการวิเคราะห์และทำนายผลให้มีประสิทธิภาพมากขึ้น เพื่อลดความเสี่ยงในการลงทุน และสร้างผลตอบแทนให้ได้สูงที่สุด

ในปัจจุบัน สำหรับในแขนงทางด้านการลงทุน การวิเคราะห์และคาดการณ์แนวโน้มราคาหลักทรัพย์นั้นจะมีด้วยกันอยู่สองวิธีหลัก คือ การวิเคราะห์ปัจจัยพื้นฐาน (Fundamental analysis) และการวิเคราะห์ทางเทคนิค (Technical analysis) ซึ่งในส่วนของการวิเคราะห์ปัจจัยพื้นฐาน จะเป็นการวิเคราะห์โดยอาศัยการดูงบการเงินของบริษัท ผลประกอบการของบริษัท และปัจจัยทางเศรษฐกิจที่เกี่ยวข้องเพื่อการคาดการณ์แนวโน้มของราคาหลักทรัพย์ในอนาคต สำหรับการวิเคราะห์ทางเทคนิค จะเป็นการวิเคราะห์โดยอาศัยการพิจารณารูปแบบราคา และตัวชี้วัดทางเทคนิค ในการคาดการณ์แนวโน้มของราคาที่จะเกิดขึ้นในอนาคต

ในส่วนของทางด้านแขนงวิทยาการคอมพิวเตอร์ เทคนิคกระบวนการเรียนรู้ของเครื่อง (Machine learning algorithms) และการเรียนรู้เชิงลึก (Deep learning algorithms) เป็นวิธีการที่นิยมนำมาใช้ในการสร้างแบบจำลองสำหรับการทำนายแนวโน้มราคาหลักทรัพย์ ซึ่งการวิเคราะห์และทำนายแนวโน้มราคาหลักทรัพย์ด้วยกระบวนการเหล่านี้มีขั้นตอนวิธีในการเก็บข้อมูลเพื่อสร้างแบบจำลองหลากหลายวิธี เช่น การใช้ข้อมูลตลาดหุ้นย้อนหลังในอดีต (Historical stock data) ซึ่งประกอบไปด้วยราคา ปริมาณซื้อขาย และตัวชี้วัดทางเทคนิค ในการใช้เป็นตัวแปรสำหรับทำนายแนวโน้มราคาหลักทรัพย์ หรือการใช้ข้อมูลจากสื่อสังคมออนไลน์ (Text data) ในการวิเคราะห์อารมณ์

และความรู้สึก (Sentiment Analysis) ในการทำนายแนวโน้มราคาหลักทรัพย์ ซึ่งเป้าหมายในการทำนายในแต่ละวิธีนั้นต่างก็มีการทำนายแนวโน้มที่ต่างกันออกไป ประกอบไปด้วยการทำนายแบบแนวโน้มรายวัน (Daily trends) แบบแนวโน้มระยะสั้น (Short-term trends) และแบบแนวโน้มระยะยาว (Long-term trends) เป็นต้น

ในการศึกษาโครงงานพิเศษนี้มีความสนใจในการพัฒนารอบวิธีการทำนายแนวโน้มราคาหลักทรัพย์แบบรายวัน (Daily trends) ผ่านเทคนิคการทำเหมืองข้อมูลโดยใช้วิธีค้นพบโมทีฟในอนุกรมเวลา (Data mining) เพื่อค้นหารูปแบบของราคาในอดีตที่ซ่อนอยู่มาใช้เป็นตัวแปรในการวิเคราะห์และสร้างวิธีการทำนายผล โดยทำการศึกษาและเปรียบเทียบกับเทคนิควิธีการพื้นฐานที่ใช้ตัวชี้วัดทางเทคนิคเป็นตัวแปรในงานที่เกี่ยวข้องส่วนใหญ่ สำหรับวิธีการทำนายผลจะทำการศึกษาโดยใช้กรอบความน่าจะเป็นมาสร้างวิธีการทำนายผ่านวิธีการวิเคราะห์ทางสถิติ (Statistical analysis) และวิธีการเรียนรู้ของเครื่อง ได้แก่ วิธีจำแนกแบบเบย์อย่างง่าย (Naïve bayes classifier) และวิธีจำแนกต้นไม้การตัดสินใจ (Decision tree classifier) เพื่อทำนายแนวโน้มการเปลี่ยนแปลงราคาหลักทรัพย์แบบรายวัน โดยมีเป้าหมายเพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพความแม่นยำของการทำนายผลด้วยเทคนิคที่แตกต่างกัน

จากผลการทดลองพบว่า วิธีการทำนายแนวโน้มราคาหลักทรัพย์ผ่านเทคนิควิธีการทำเหมืองข้อมูลมีค่า Mean Accuracy สูงกว่าเทคนิควิธีการพื้นฐานโดยเฉลี่ยอยู่ที่ 1-3% และค่า Mean Precision สูงกว่าเทคนิควิธีการพื้นฐานโดยเฉลี่ยอยู่ที่ 1-8% โดยสำหรับเทคนิควิธีการทำเหมืองข้อมูลวิธีการจำแนกแบบเบย์อย่างง่าย (Naïve bayes classifier) เป็นวิธีที่ให้ค่า Accuracy และ Precision เฉลี่ยสูงสุด คือ 0.6022 และ 0.6347 ตามลำดับ และในส่วนของเทคนิควิธีการพื้นฐาน วิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multilayer perceptron classifier) เป็นวิธีที่ให้ค่า Accuracy เฉลี่ยสูงสุด คือ 0.5725 และวิธีจำแนกต้นไม้การตัดสินใจ (Decision tree classifier) เป็นวิธีที่ให้ค่า Precision เฉลี่ยสูงสุด คือ 0.5995 ซึ่งเมื่อพิจารณาโดยรวมจะเห็นได้ว่าวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูลที่ได้ทดลองศึกษาสามารถแสดงค่าความถูกต้องและค่าความแม่นยำได้สูงขึ้นเมื่อเทียบกับวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค

1.2 วัตถุประสงค์

โครงการพิเศษนี้มีเป้าหมายเพื่อทดลองศึกษาและเปรียบเทียบวิธีการทำนายแนวโน้มราคาหลักทรัพย์ผ่านเทคนิคการทำเหมืองข้อมูลและเทคนิควิธีการพื้นฐาน เพื่อให้บรรลุเป้าหมายดังกล่าว จึงกำหนดวัตถุประสงค์ของโครงการดังต่อไปนี้

1. เพื่อศึกษาและพัฒนาวิธีการทำนายการเปลี่ยนแปลงของแนวโน้มราคาหลักทรัพย์ด้วยเทคนิคการทำเหมืองข้อมูลผ่านกรอบความน่าจะเป็นด้วยการวิเคราะห์ทางสถิติ และวิธีการเรียนรู้ของเครื่อง ได้แก่ ตัวจำแนกแบบเบย์อย่างง่าย และตัวจำแนกต้นไม้การตัดสินใจ
2. เพื่อศึกษาและเปรียบเทียบประสิทธิภาพความแม่นยำของวิธีการทำนายผลด้วยเทคนิคการทำเหมืองข้อมูลโดยใช้วิธีการค้นพบโมทีฟในอนุกรมเวลา และเทคนิควิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค
3. เพื่อศึกษาวิเคราะห์และตัดสินใจซื้อขายหลักทรัพย์ผ่านการวิเคราะห์ทางสถิติและผลการทำนายแนวโน้มจากโมเดลแบบรายวัน

1.3 ขอบเขตของโครงการ

1. ชุดข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลตลาดหุ้นจากตลาดหลักทรัพย์แห่งประเทศไทย (SET) ในหลากหลายอุตสาหกรรมที่แตกต่างกันแบบสุ่มทั้งหมด 5 รายการ ได้แก่ KBANK (กลุ่มธุรกิจการเงิน), PTTEP (กลุ่มทรัพยากร), CPALL (กลุ่มบริการ), DELTA (กลุ่มเทคโนโลยี) และ SCC (กลุ่มอสังหาริมทรัพย์และก่อสร้าง)
2. ข้อมูลตัวแปรที่นำมาศึกษา ได้แก่ ราคาปิด (Close) และปริมาณซื้อขาย (Volume) ของชุดข้อมูลในทุกรายการ
3. วิธีการทำนายผลที่ใช้ในการศึกษาผ่านเทคนิควิธีการทำเหมืองข้อมูล ได้แก่ วิธีการวิเคราะห์ทางสถิติ วิธีจำแนกแบบเบย์อย่างง่าย และวิธีจำแนกต้นไม้การตัดสินใจ

1.4 ประโยชน์ของโครงการ

1. เป็นแนวทางในการศึกษาแก่ผู้ที่สนใจศึกษาเทคนิควิธีการต่างๆในการดึงคุณลักษณะในการใช้วิเคราะห์และทำนายแนวโน้มราคาหลักทรัพย์
2. เพื่อพัฒนาวิธีการทำนายผลด้วยเทคนิคการทำเหมืองข้อมูลโดยเปรียบเทียบกับเทคนิควิธีการพื้นฐาน
3. สามารถนำไปทดลองศึกษาและประยุกต์ใช้กับข้อมูลหลักทรัพย์อื่นๆที่มีความผันผวนสูงขึ้น เช่น ตลาดสกุลเงินดิจิทัล หรือข้อมูลอนุกรมเวลาประเภทอื่นๆ เป็นต้น

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดทฤษฎีที่เกี่ยวข้อง

2.1.1 Domain Knowledge

2.1.1.1 วิธีการทำนายแนวโน้มตลาดหลักทรัพย์ในปัจจุบัน (Stock Market Forecasting)

(1) การวิเคราะห์ปัจจัยพื้นฐาน (Fundamental Analysis)

การวิเคราะห์ปัจจัยพื้นฐาน (Fundamental Analysis) เป็นวิธีการวิเคราะห์หลักทรัพย์โดยจะมุ่งเน้นไปที่การประเมินมูลค่าที่แท้จริงของหลักทรัพย์ (Intrinsic Value) ซึ่งในปัจจุบันนักลงทุนสามารถใช้ผลหรือข้อมูลจากการวิเคราะห์ดังกล่าว เป็นเกณฑ์ประกอบการพิจารณาตัดสินใจลงทุนได้ เช่น จะซื้อหลักทรัพย์ก็ต่อเมื่อราคาตลาดของหลักทรัพย์นั้นต่ำกว่ามูลค่าตามปัจจัยพื้นฐานที่คำนวณได้ และจะทำการขายหลักทรัพย์เมื่อพบว่าราคาตลาดของหลักทรัพย์นั้นสูงกว่ามูลค่าตามราคาพื้นฐาน ซึ่งการวิเคราะห์หลักทรัพย์โดยใช้ปัจจัยพื้นฐาน จะเป็นการวิเคราะห์ถึงภาวะและแนวโน้มทางเศรษฐกิจโดยทั่วไป ซึ่งอาจพิจารณาถึงเศรษฐกิจโลกที่มีผลกระทบต่อประเทศ รวมถึงนโยบายเศรษฐกิจของรัฐบาล นอกจากนี้ยังวิเคราะห์ถึงภาวะอุตสาหกรรมที่เกี่ยวข้อง และตัวบริษัทผู้ออกหลักทรัพย์ด้วย ซึ่งจะเป็นการวิเคราะห์ทั้งในเชิงคุณภาพ เช่น ประสิทธิภาพการบริหารของผู้บริหาร เป็นต้น และการวิเคราะห์ในเชิงปริมาณ เช่น ผลการดำเนินงาน และฐานะทางการเงินของบริษัท รวมทั้งอัตราส่วนทางการเงินต่างๆ เป็นต้น [1]

(2) การวิเคราะห์ทางเทคนิค (Technical Analysis)

การวิเคราะห์ทางเทคนิค (Technical Analysis) คือ การศึกษาแบบราคาของสินทรัพย์ทางการเงินในสินทรัพย์ใดๆ เช่น หลักทรัพย์ ฟอเร็กซ์ โดยมีเป้าหมายเพื่ออธิบายเหตุผลของการกำหนดรูปแบบราคาเหล่านั้น ซึ่ง

เป็นการนำไปสู่การวิเคราะห์และประเมินทิศทางความเป็นไปได้ของราคา โดยทั่วไปสามารถแบ่งออกได้เป็น 3 ลักษณะ ได้แก่

1) การวิเคราะห์โดยใช้รูปแบบกราฟแท่งเทียน (Candlestick Patterns)

การวิเคราะห์ในรูปแบบนี้ เทรดเดอร์หรือนักเก็งกำไร จะใช้กราฟแท่งเทียนโดยพิจารณาที่บริเวณแนวรับ-แนวต้าน ในปัจจุบันเทคนิคนี้ได้รับความนิยมสูง เนื่องจากมีความซับซ้อนน้อย และไม่ได้ลดทอนประสิทธิภาพในการวิเคราะห์ผลแต่อย่างใด โดยจะทำการพิจารณาเพียง ราคาเปิด (Open) ราคาปิด (Close) ราคาสูงสุด (High) และราคาต่ำสุด (Low) ของแท่งเทียน (Candlestick) หนึ่งๆ เท่านั้น

2) การวิเคราะห์โดยใช้รูปแบบกราฟราคา (Chart Patterns)

การวิเคราะห์ในรูปแบบนี้ จะเป็นการพิจารณาชุดของแท่งเทียน ที่ประกอบรวมกันในช่วงเวลาหนึ่งๆ โดยเทรดเดอร์หรือนักเก็งกำไร จะใช้เส้นแนวโน้ม (Trendline) เป็นเครื่องมือหลักในการกำหนดรูปแบบที่เป็นไปได้ เช่น รูปแบบกราฟสามเหลี่ยมสมมาตร (Symmetrical Triangle) เป็นต้น

3) การวิเคราะห์โดยใช้การใช้อินดิเคเตอร์ (Indicators)

การวิเคราะห์โดยใช้แนวทางในการวิเคราะห์เชิงเทคนิคส่วนใหญ่ จะอ้างอิงกับค่าเฉลี่ยของราคา ซึ่งอาจนำค่าเฉลี่ยมาพล็อตกราฟโดยตรง เช่น Moving Average (MA) หรืออาจนำราคามาแปลงเป็นค่าดัชนี เช่น เครื่องมืออย่าง RSI และ Stochastic เป็นต้น [2]

2.1.1.2 ตัวชี้วัดทางเทคนิค (Technical Indicators)

ตัวชี้วัดทางเทคนิค (Technical Indicators) เป็นเครื่องมือที่ถูกคิดค้นขึ้นมาเพื่อช่วยในการวิเคราะห์ทางเทคนิคสำหรับการซื้อขาย ไม่ว่าจะเป็นหลักทรัพย์ อนุพันธ์ ฟอเร็กซ์ หรือแม้แต่สกุลเงินดิจิทัล ซึ่งอินดิเคเตอร์จะเป็นการคำนวณโดยใช้ราคาหุ้นมาเข้าสู่สูตรทางคณิตศาสตร์ โดยในอินดิเคเตอร์แต่ละตัวก็จะมีวิธีการคำนวณที่ต่างกันออกไป อินดิเคเตอร์แบ่งออกได้เป็น 4 ประเภทตามวัตถุประสงค์การใช้งาน ดังนี้ [3]

1) Trend Indicator คือ อินดิเคเตอร์ที่ใช้ในการระบุทิศทางของแนวโน้มว่าเป็นช่วงขาขึ้น (Uptrend) ขาลง (Downtrend) หรือออกข้าง (Sideway) ยกตัวอย่างเช่น Moving Average (MA), Bollinger Bands (BB) และ Parabolic SAR เป็นต้น

2) Volume Indicators คือ อินดิเคเตอร์ที่ใช้ในการวิเคราะห์ปริมาณการซื้อขาย (Trading Volume) ซึ่งมีบทบาทที่สำคัญในการวิเคราะห์ทางเทคนิค เนื่องจากช่วยในการยืนยันแนวโน้มและรูปแบบของราคาได้ และยังสามารถแสดงให้เห็นถึงความสนใจของนักลงทุนในตลาดว่าสินทรัพย์ไหนที่กำลังเป็นที่สนใจหรือมีแนวโน้ม ยกตัวอย่างเช่น Volumes (Vol) และ On Balance Volume (OBV) เป็นต้น

3) Momentum Indicators คือ อินดิเคเตอร์ที่ใช้ในการวิเคราะห์แรงเหวี่ยง หรืออัตราเร่งของราคา ซึ่งสามารถแสดงให้เห็นถึงความแข็งแกร่งของการเคลื่อนไหวของราคาได้ อย่างในบางครั้งราคาอาจร่วงลงอย่างรวดเร็ว หรือบางครั้งอาจร่วงลงช้า เราสามารถวิเคราะห์ความเร็วของการลดลงหรือเพิ่มขึ้นของราคาจากแรงเหวี่ยงหรืออัตราเร่งได้ ยกตัวอย่างเช่น Commodity channel index (CCI), Relative Strength Index (RSI) และ Stochastic เป็นต้น

4) Volatility Indicator คือ อินดิเคเตอร์ที่ใช้ในการวิเคราะห์ความผันผวนของราคา ซึ่งความผันผวนต่ำ หมายถึงมีการเคลื่อนไหวของราคาที่น้อย ในขณะที่ความผันผวนสูง หมายถึงมีการเคลื่อนไหวของราคาที่รุนแรง ซึ่งมักจะคาดเดาได้ยากในระยะสั้น ยกตัวอย่างเช่น Average True Range (ATR) และ Standard Deviation (SD) เป็นต้น [4]

2.1.2 Computer Science Knowledge

2.1.2.1 การเตรียมข้อมูล (Data Preprocessing)

การเตรียมข้อมูลเป็นขั้นตอนหลังจากการทำความสะอาดข้อมูล (Data Cleaning) ซึ่งเป็นการเตรียมข้อมูลให้พร้อมนำไปใช้ในการสร้างโมเดลหรือนำไปประมวลผลในขั้นถัดไป เช่น การคัดเลือกฟีเจอร์ (Feature selection) และการแปลงข้อความเป็นตัวเลข (Encoding) เป็นต้น ซึ่งในขั้นตอนนี้จะทำให้ชุดข้อมูลดิบ (Raw Data) ที่ได้มาในขั้นตอนแรก มีความถูกต้องแม่นยำและพร้อมในการนำไปประมวลผลในขั้นตอนต่อไป [5]

(1) การปรับขนาดข้อมูล (Data Normalization)

2.1.2.1.1.1 Min-Max Normalization

เป็นวิธีการปรับขนาดของข้อมูล ให้อยู่ในช่วง $[0, 1]$ โดยมีสมการดังนี้

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

โดยที่ x คือ ข้อมูลที่อยู่ในตัวแปรหรือคอลัมน์

$\min(x)$ คือ ค่าที่น้อยที่สุด (Minimum) ที่อยู่ในคอลัมน์

$\max(x)$ คือ ค่าที่มากที่สุด (Maximum) ที่อยู่ในคอลัมน์

2.1.2.1.1.2 Standard Normalization

เป็นวิธีการปรับให้ข้อมูลมีค่าเฉลี่ย (Mean) เท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เท่ากับ 1 โดยมีสมการดังนี้

$$x_{std} = \frac{x - \text{mean}(x)}{\text{std}(x)}$$

โดยที่ x คือ ข้อมูลที่อยู่ในตัวแปรหรือคอลัมน์

$\text{mean}(x)$ คือ ค่าเฉลี่ยของข้อมูลทั้งหมดที่อยู่ในคอลัมน์

$\text{std}(x)$ คือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูลทั้งหมดที่อยู่ในคอลัมน์ [6]

(2) การแปลงข้อมูล (Data Transformation)

การแปลงข้อมูลเป็นขั้นตอนในการแปลงรูปแบบ โครงสร้าง หรือค่าของข้อมูล ให้อยู่ในรูปแบบที่เหมาะสมและพร้อมนำไปใช้งาน [7]

2.1.2.2 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data mining) เป็นกระบวนการในการค้นหารูปแบบในชุดข้อมูลขนาดใหญ่ โดยใช้วิธีการของการเรียนรู้ของเครื่อง สถิติ และระบบฐานข้อมูล [8][9] การทำเหมืองข้อมูลเป็นขั้นตอนวิธีการในการค้นหาความรู้ในฐานข้อมูล โดยการทำเหมืองข้อมูลเป็นเทคนิคเพื่อค้นหารูปแบบ (Pattern) จากข้อมูลจำนวนมากโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ หรือในอีกนัยหนึ่งการทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

(1) เทคนิคการค้นพบโมทีฟและดิสคอร์ด (Motif and Discord Discoveries)

2.1.2.2.1 การค้นพบโมทีฟ (Motif Discovery)

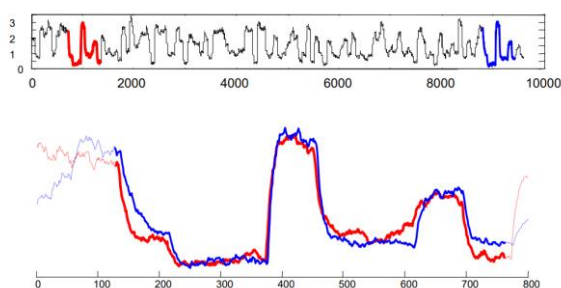


Figure 1: (top) The output steam flow telemetry of the Steamgen dataset has a motif of length 640 beginning at locations 589 and 8,895. (bottom) by overlaying the two motifs we can see how remarkably similar they are to each other

ภาพที่ 2.1 ตัวอย่างวิธีการค้นพบโมทีฟในข้อมูลอนุกรมเวลา

การค้นพบโมทีฟเป็นการค้นหาของลำดับย่อยอนุกรมเวลาในชุดข้อมูลอนุกรมเวลาที่มีลักษณะรูปแบบที่มีลักษณะคล้ายคลึงกันมากที่สุด ซึ่ง

เทคนิควิธีนี้ได้ถูกนำไปใช้ในหลากหลายโดเมน ได้แก่ ทางการแพทย์ ทางชีววิทยา และการพยากรณ์สภาพอากาศ เป็นต้น [10]

2.1.2.2.1.2 การค้นพบดิสคอร์ด (Discord Discovery)

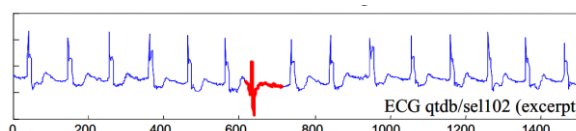


Figure 1: The time series discord found in an excerpt of electrocardiogram qtdb/sel102 (marked in bold line). The location of the discord exactly coincides with a premature ventricular contraction

ภาพที่ 2.2 ตัวอย่างวิธีการค้นพบดิสคอร์ดในข้อมูลอนุกรมเวลา

การค้นพบดิสคอร์ดเป็นการค้นหาลำดับย่อยอนุกรมเวลาในชุดข้อมูลอนุกรมเวลาที่มีลักษณะรูปแบบที่แตกต่างจากลำดับย่อยอื่นๆ ในชุดข้อมูลอนุกรมเวลาทั้งหมด ซึ่งสามารถใช้ตรวจจับลักษณะของลำดับย่อยที่ผิดปกติได้ ซึ่งเทคนิควิธีนี้ส่วนใหญ่ใช้ในขั้นตอนการทำเหมืองข้อมูล เช่น การปรับปรุงประสิทธิภาพในการจัดกลุ่มข้อมูล (Clustering) การทำความสะอาดข้อมูล (Data cleaning) และการตรวจจับสิ่งผิดปกติ (Anomaly detection) เป็นต้น [11]

(2) การวัดระยะห่างความคล้ายคลึงกัน (Similarity Distance Measure)

ในเทคนิควิธีการค้นพบโมทีฟและดิสคอร์ดของข้อมูลอนุกรมเวลาที่ต้องมีการวัดความคล้ายคลึงกันของแต่ละลำดับย่อยอนุกรมเวลาเพื่อหาคู่ลำดับที่คล้ายคลึงกันหรือต่างกันมากที่สุด จะต้องมีการใช้มาตรวัดระยะห่างระหว่างคู่ลำดับย่อย โดยมาตรวัดที่นิยมใช้ในการวัดระยะห่างของข้อมูลอนุกรมเวลาส่วนใหญ่ ได้แก่ มาตรวัดระยะทางแบบยูคลิด (Euclidean Distance) และมาตรวัดระยะทางแบบไดนามิกไทม์วอร์ปิง (Dynamic Time Warping)

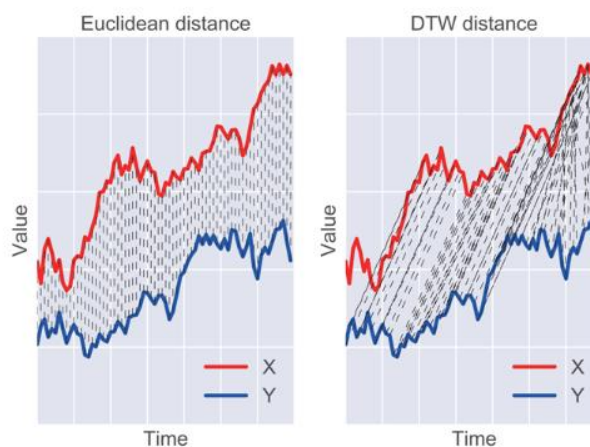


FIGURE 1 Correspondence of time-series data [Color figure can be viewed at wileyonlinelibrary.com]

ภาพที่ 2.3 เปรียบเทียบการวัดระยะห่างระหว่าง Euclidean Distance กับ Dynamic Time Warping

2.1.2.2.2.1 การวัดระยะห่างแบบยุคลิด (Euclidean Distance)

การวัดระยะห่างแบบยุคลิด (Euclidean Distance) เป็นวิธีการวัดระยะห่างที่ง่ายที่สุดจากวิธีในการคำนวณระยะห่างทั้งหมด โดยเป็นวิธีการวัดระยะทางระหว่างจุดสองจุด ซึ่งสามารถคำนวณได้จาก

$$D(X(t), Y(t)) = \sum_{i=1}^N \sqrt{|X_i - Y_i|^2}$$

ซึ่งข้อเสียหลักของวิธีวัดระยะห่างแบบยุคลิด คือ ไม่สามารถวัดระยะห่างของข้อมูลอนุกรมเวลาที่มีจำนวนระยะเวลาในการเก็บไม่เท่ากัน

2.1.2.2.2.2 การวัดระยะทางแบบไดนามิกไทม์วอร์ปิง (Dynamic Time Warping)

การวัดระยะทางแบบไดนามิกไทม์วอร์ปิง (Dynamic Time Warping) เป็นวิธีการกำหนดพลวัต (dynamic programming) ที่มีความแม่นยำสูง ซึ่งสามารถใช้วัดความต่างระหว่างข้อมูลอนุกรมเวลาสองชุดที่มี

ระยะเวลาในการเก็บไม่เท่ากันได้ โดยกำหนดให้ข้อมูลอนุกรมเวลา 2 ชุด
เป็นดังนี้

$X(t) = X_1, X_2, \dots, X_e, \dots, X_N$ ขนาด N ระยะเวลา

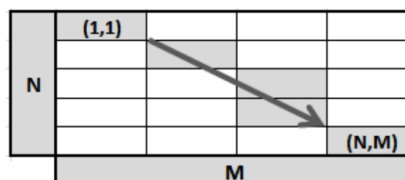
$Y(t) = Y_1, Y_2, \dots, Y_f, \dots, Y_M$ ขนาด M ระยะเวลา

โดยในส่วนของการวิธีการคำนวณระยะทาง มีขั้นตอนดังนี้

- 1) สร้างเมทริกซ์ขนาด $N \times M$ มิติ เมื่อส่วนประกอบที่ (e^{th}, f^{th}) ของเมทริกซ์ คือ ระยะห่างกำลังสอง (squared distance) ของ X_e และ Y_f

$$d(X_e, Y_f) = (X_e - Y_f)^2$$

- 2) หาเส้นทางการวอร์ป (Warping path) ทั้งหมดที่เป็นไปได้ โดยเส้นทางการวอร์ป คือ เส้นทางการเคลื่อนที่จาก $(1,1)$ ถึง (N,M) ซึ่งมีได้หลายรูปแบบ ดังแสดงในรูปที่ 1



ภาพที่ 2.4 เส้นทางการวอร์ป (warping path)

- 3) ดังนั้นเส้นทางที่ดีที่สุด คือ เส้นทางที่มีต้นทุนการวอร์ป (warping cost) ต่ำที่สุดเมื่อ เปรียบเทียบกับเส้นทางการวอร์ป ทั้งหมดที่เป็นไปได้ การวัดระยะทาง Dynamic Time Warping ที่มีต้นทุน ต่ำที่สุดระหว่าง $X(t)$ และ $Y(t)$ สามารถคำนวณดังนี้

$$D(X(t), Y(t)) = \min \sqrt{\sum_{g=1}^G w_g}$$

โดย w_g คือ องค์ประกอบของเมทริกซ์เส้นทางวอร์ป $(i, j)_g$ ในรูปที่ 1 และเป็นองค์ประกอบที่ g ของ เส้นทางการวอร์ป (warping path) จนถึง (N, M) และ G คือ จุดสุดท้ายของเส้นทาง [12]

โดยอัลกอริทึมการวัดระยะทางแบบไดนามิกไทม์วอร์ปิง (Dynamic Time Warping) สามารถแสดงเป็น Pseudocode ได้ดังนี้ [13]

ALGORITHM 1 DTW distance

```

1: Procedure DTW( $x, y, w = 5$ )
                                     > Initialize Matrix D
2:   Var  $D[N, M]$ 
3:    $D[1, 1] = 0$ 
4:   for  $i = 2$  to  $N$  do
5:     for  $j = 2$  to  $M$  do
6:        $D[i, j] = \infty$ 
7:     end for
8:   end for
                                     > Calculate DTW distance
9:   for  $i = 2$  to  $N$  do
10:    for  $j = \max(1, i - w)$  to  $\min(M, i + w)$  do
11:       $D[i, j] = d(x[i - 1], y[j - 1])$ 
         $+ \min(D[i, j - 1], D[i - 1, j], D[i - 1, j - 1])$ 
12:    end for
13:  end for
14:  return  $D[N, M]$ 
15: end procedure

```

ภาพที่ 2.5 Pseudocode แสดงขั้นตอนวิธีการทำงานของมาตรวัดแบบ

Dynamic Time Warping

2.1.2.3 วิธีการทำนายผล (Predictive Modeling)

วิธีการทำนายผลเป็นกระบวนการเชิงคณิตศาสตร์ที่ใช้ในการทำนายเหตุการณ์หรือผลลัพธ์ที่อาจเกิดขึ้นในอนาคตโดยอาศัยการวิเคราะห์รูปแบบหรือความสัมพันธ์จากข้อมูลที่มี โดยวิธีการทำนายผลเป็นส่วนหนึ่งที่สำคัญของการวิเคราะห์เชิงคาดการณ์ (Predictive Analytics) ซึ่งเป็นลักษณะการวิเคราะห์ข้อมูลโดยใช้ทั้งข้อมูลในอดีตและปัจจุบันในการทำนายเหตุการณ์ กิจกรรม พฤติกรรม หรือแนวโน้มที่คาดว่าจะเกิดขึ้น เป็นต้น [14]

(1) วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

การวิเคราะห์ทางสถิติเป็นกระบวนการเก็บรวบรวมและอธิบายข้อมูลที่มีเพื่อค้นหารูปแบบและแนวโน้มที่ซ่อนอยู่ ซึ่งเป็นส่วนหนึ่งของกระบวนการวิเคราะห์ข้อมูล โดยการวิเคราะห์ทางสถิติสามารถประยุกต์ใช้ได้ในการรวบรวมการตีความ การวิจัย การสร้างแบบจำลองเชิงสถิติ หรือการออกแบบการสำรวจ เป็นต้น [15]

2.1.2.3.1.1 การทดลองสุ่ม ปรัชมิตัวอย่าง และเหตุการณ์

2.1.2.3.1.1.1 การทดลองสุ่ม (Random Experiment)

การทดลองสุ่ม (Random Experiment) คือ การทดลองใดๆ ที่มีผลลัพธ์ (Outcome) ที่เป็นไปได้มากกว่าหนึ่งอย่าง และไม่สามารถทำนายผลลัพธ์ที่แน่นอนได้ล่วงหน้า แต่จะทราบผลลัพธ์ที่เป็นไปได้ทั้งหมด

2.1.2.3.1.1.2 ปรัชมิตัวอย่าง (Sample Space)

ปรัชมิตัวอย่าง (Sample Space) คือ เซตของผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองสุ่มใดๆ

2.1.2.3.1.1.3 เหตุการณ์ (Event)

เหตุการณ์ (Event) คือ เซตย่อยที่เราสนใจของผลลัพธ์ทั้งหมดจากการทดลองสุ่ม หรือเป็นเซตย่อยของปรัชมิตัวอย่าง

2.1.2.3.1.2 การหาค่าความน่าจะเป็นของเหตุการณ์ (Probability)

ความน่าจะเป็น (Probability) ของเหตุการณ์ใดๆ คือ ค่าที่ใช้วัดโอกาสของการเกิดเหตุการณ์ที่เราสนใจนั้นว่ามีมากน้อยเพียงใด และค่าความน่าจะเป็นมีค่าตั้งแต่ 0 ถึง 1 ถ้าความน่าจะเป็นของเหตุการณ์มีค่าเข้าใกล้ 1 แสดงว่าเหตุการณ์นั้นมีโอกาสเกิดขึ้นค่อนข้างมาก และถ้าความน่าจะเป็นของเหตุการณ์มีค่าเข้าใกล้ 0 ก็แสดงว่าเหตุการณ์นั้นมีโอกาสเกิดขึ้นได้น้อย

$$P(E) = \frac{n(E)}{n(S)}$$

โดยที่ $P(E)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ใดๆ ที่เราสนใจ

$n(E)$ คือ จำนวนผลลัพธ์ของเหตุการณ์ที่เราสนใจ

$n(S)$ คือ จำนวนผลลัพธ์ทั้งหมดที่อาจจะเกิดขึ้นได้

2.1.2.3.1.3 การกำหนดตัวแปรสุ่ม (Random Variable)

ตัวแปรสุ่ม (Random Variable) คือ ฟังก์ชันที่มีค่าเป็นเลขจำนวนจริง ซึ่งเกิดจากสมาชิกแต่ละสมาชิกในปริภูมิตัวอย่างของผลลัพธ์ทั้งหมดในการทดลองเชิงสุ่มใดๆ

2.1.2.3.1.3.1 การหาค่าความคาดหวังของตัวแปรสุ่ม (Expectation of Random Variable)

ค่าคาดหวังของตัวแปรสุ่ม คือ ค่าเฉลี่ยของตัวแปรสุ่ม หรือค่าคาดคะเนของตัวแปรสุ่ม ซึ่งต้องอาศัยทฤษฎีความน่าจะเป็นมาใช้ โดยมีวิธีการคำนวณดังนี้

$$E[X] = \sum x_i p(x_i)$$

โดยที่ x_i คือ ตัวแปรสุ่มลำดับที่ i

$p(x_i)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ x_i

2.1.2.3.1.3.2 การหาค่าความแปรปรวนของตัวแปรสุ่ม (Variance of Random Variable)

ความแปรปรวนของตัวแปรสุ่มของ X เป็นค่าที่บอกให้ทราบถึงการกระจายของตัวแปรสุ่ม X ว่ามีความแตกต่างไปจากค่าเฉลี่ยหรือค่าคาดหวังของตัวแปรสุ่มนั้นมากน้อยเพียงใด โดยมีวิธีการคำนวณดังนี้

$$\text{var}(X) = E[(X - E[X])^2]$$

โดยที่ X คือ ตัวแปรสุ่ม X

$E[X]$ คือ ค่าเฉลี่ยของตัวแปรสุ่ม X

(2) วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier) เป็นหนึ่งในโมเดลของวิธีการเรียนรู้ของเครื่อง (Machine Learning) ที่พัฒนาขึ้นมาจากทฤษฎีความน่าจะเป็นของเบย์ (Bayes Theorem) ซึ่งเป็นโมเดลที่จัดอยู่ในกลุ่ม Supervised Learning โดยเป็นการทำผลแบบการจำแนกประเภท (Classification)

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

โดยที่ c คือ คลาส (class)

x คือ ตัวแปร หรือคุณลักษณะ (Features)

P คือ ความน่าจะเป็นของเหตุการณ์ (Probability)

$P(c|x)$: *Posterior Probability* คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ในคลาส c เมื่อข้อมูลมีตัวแปรหรือคุณลักษณะ x

$P(x|c)$: *Likelihood* คือ ความน่าจะเป็นที่ข้อมูลจะมีตัวแปรหรือคุณลักษณะ x เมื่อทราบเหตุการณ์ในคลาส c

$P(c)$: *Class Prior Probability* คือ ความน่าจะเป็นของแต่ละเหตุการณ์ในคลาส c จากจำนวนเหตุการณ์ทั้งหมด

$P(x)$: *Predictor Prior Probability* คือ ความน่าจะเป็นของตัวแปรหรือคุณลักษณะ x

ข้อดีของวิธีจำแนกแบบเบย์อย่างง่าย (Advantage of Naïve Bayes)

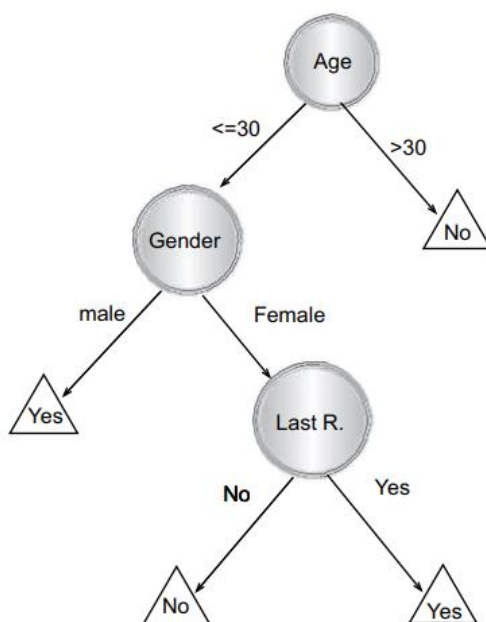
- 1) เนื่องจากโมเดล Naive Bayes ใช้หลักการบนพื้นฐานของความน่าจะเป็น ดังนั้น จึงเรียนรู้ง่ายและเข้าใจได้เร็ว
- 2) ใช้ทรัพยากรในการประมวลผลน้อยกว่า และให้ผลลัพธ์ออกมาอย่างรวดเร็ว
- 3) มีความยืดหยุ่นต่อการเพิ่มหรือลดจำนวนตัวอย่างข้อมูล
- 4) รองรับทั้งข้อมูลที่มีค่าต่อเนื่อง (Continuous) และไม่ต่อเนื่อง (Discrete)
- 5) ไม่เกิดปัญหา Overfitting เหมือนที่มีในโมเดลอื่นๆ
- 6) จัดการข้อมูลที่สูญหาย (Missing Value) ได้ง่ายกว่า
- 7) รองรับการทำ Train Model สำหรับข้อมูลตัวอย่างที่มีปริมาณมากๆ ได้ดี

ข้อเสียของวิธีจำแนกแบบเบย์อย่างง่าย (Disadvantage of Naïve Bayes)

- 1) ไม่รองรับการทำนายผลแบบ Regression หรือผลลัพธ์ที่ออกมาเป็นตัวเลข
- 2) กรณีที่เราใช้โมเดล Naive Bayes แบบการนับจำนวนหรือความถี่สำหรับคำนวณหาความน่าจะเป็น หากข้อมูลที่เราจะทำนายผลไม่ปรากฏอยู่ในข้อมูลตัวอย่างมาก่อน ความถี่จะมีค่าเป็น 0 (Zero Frequency) ดังนั้น ความน่าจะเป็นที่คำนวณได้ก็จะเท่ากับ 0 เช่นกัน ซึ่งอาจเกิดข้อผิดพลาดในการทำนายผล
- 3) โมเดล Naive Bayes เหมาะกับการทำนายผลกับข้อมูลที่มีลักษณะการแบ่งกลุ่ม (Category) ซึ่งข้อมูลที่เป็นเช่นนี้ในการใช้งานจริง (Real World) อาจมีไม่มากนัก

(3) วิธีจำแนกไม้การตัดสินใจ (Decision Tree Classifier)

ต้นไม้การตัดสินใจ (Decision Tree) เป็นหนึ่งในโมเดลที่สำคัญของวิธีการเรียนรู้ของเครื่อง (Machine Learning) โดยโมเดลนี้จะมีการสร้างลำดับขั้นตอนการตัดสินใจเป็นแผนภูมิที่มีลักษณะโครงสร้างคล้ายกับต้นไม้แบบกลับหัว ซึ่งโมเดลจะมีการเรียนรู้ในการจำแนกข้อมูลออกเป็นกลุ่มต่างๆโดยใช้ตัวแปรหรือคุณลักษณะในการจำแนกประเภทข้อมูล



ภาพที่ 2.6 ตัวอย่างแผนภูมิต้นไม้การตัดสินใจ

องค์ประกอบของโครงสร้างต้นไม้การตัดสินใจ

- 1) **Node** เป็น จุดที่ใช้กำหนดเงื่อนไข หรือเทียบได้กับปมของต้นไม้
- 2) **Branch** คือ การแบ่งเงื่อนไขออกเป็นกรณีย่อยๆ หรือเทียบได้กับการแตกกิ่งของต้นไม้
- 3) **Root Node** คือ Node ชั้นบนสุด หรือเงื่อนไขแรกของแผนภูมิ ซึ่งเทียบได้กับรากของต้นไม้ (แบบกลับหัว)

- 4) **Internal Node (Decision Node)** คือ Node ทั้งหมดในลำดับชั้นต่อจาก Root Node
- 5) **Leaf Node (Terminal Node)** คือ Node ในลำดับชั้นล่างสุด หรือ Node ลำดับสุดท้ายของแผนภูมิ และจะไม่มีแตกกิ่ง (แบ่งเงื่อนไข) ต่อไปอีก ซึ่งก็เทียบได้กับใบของต้นไม้

Gini Index

Gini Index เป็นฟังก์ชันที่ใช้ในการแยกการปะปนกันของข้อมูล โดยสามารถคำนวณได้จาก

$$Gini(S) = 1 - \sum_{i=1}^N [P_i]^2$$

โดยที่ N คือ ประเภทของคลาส (Class) ทั้งหมด

P_i คือ สัดส่วนของข้อมูลในแต่ละคลาส

Gain

Gain เป็นฟังก์ชันในการเลือกฟีเจอร์ (features) มาสร้างเป็น Node ในต้นไม้ โดยฟังก์ชัน Gain ที่ใช้ Gini ในการเลือกฟีเจอร์สามารถคำนวณได้จาก

$$Gain(S, A) = Gini(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} (1 - \sum_{i=1}^c P_{iv})$$

โดยที่ S คือ ข้อมูลทั้งหมด

A คือ ตัวแปร (Features)

$value(A)$ คือ ประเภทในตัวแปร

V คือ สมาชิกประเภทในตัวแปร

S_v คือ ข้อมูลประเภท V ที่อยู่ในตัวแปร A

ข้อดีของโมเดลต้นไม้การตัดสินใจ (Advantage of Decision Tree)

- 1) สามารถเรียนรู้และทำความเข้าใจได้ง่าย
- 2) รองรับการทำนายผลทั้งแบบ Classification และ Regression
- 3) ไม่เกิดปัญหาเมื่อข้อมูลมีค่าที่สูญหาย (Missing Value)
- 4) สามารถแสดงผลในรูปแบบ Visual ได้

ข้อเสียของโมเดลต้นไม้การตัดสินใจ (Disadvantage of Decision Tree)

- 1) มีโอกาสเกิดปัญหา Overfitting ได้
- 2) ใช้เวลาในการประมวลผลค่อนข้างนาน [16]

2.1.2.4 วิธีการวัดประสิทธิภาพ (Evaluation Metrics)

วิธีการวัดประสิทธิภาพ เป็นการวัดคุณภาพของโมเดลทางสถิติ หรือวิธีการเรียนรู้ของเครื่อง โดยการใช้มาตรวัดประสิทธิภาพนั้นเป็นขั้นตอนที่สำคัญอย่างมากในการประเมินการทำงานของโมเดลว่าทำงานได้อย่างถูกต้องและเหมาะสมหรือไม่ [17]

(1) คะแนน Accuracy

Accuracy Score คือค่าความแม่นยำในการทำนายผล โดยคิดจากจำนวนที่ทายถูกทั้งหมดหารด้วยจำนวนรายการทั้งหมดใน Confusion Matrix โดยมีวิธีการคำนวณดังนี้

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(2) คะแนน Precision

Precision Score คือ อัตราส่วนของการทำนายเชิงบวกแล้วถูกต้อง (TP) ต่อ จำนวนการทำนายเชิงบวกทั้งหมด โดยมีวิธีการคำนวณดังนี้

$$Precision = \frac{TP}{TP + FP}$$

(3) คะแนน Recall

Recall Score คือ อัตราส่วนของการทำนายเชิงบวกแล้วถูกต้อง (TP) ต่อผลลัพธ์จริงในเชิงบวก (Actual Positive) ทั้งหมด โดยมีวิธีการคำนวณดังนี้

$$Recall = \frac{TP}{TP + FN}$$

(4) คะแนน F1

F1-Score คือ การวัดความสามารถของโมเดลแบบ Single metric ซึ่งเป็น การคำนวณเพื่อหาค่าเฉลี่ย (Mean) ระหว่างค่า Precision และ Recall โดยวิธีการ คำนวณดังนี้ [18]

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

2.2 งานวิจัยที่เกี่ยวข้อง

Artificial Intelligence Applied to Stock Market Trading: A Review (2021)

FERNANDO G. D. C. FERREIRA, AMIR H. GANDOMI และ RODRIGO T. N. CARDOSO ได้นำเสนอการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องในการนำปัญญาประดิษฐ์ (AI) มาประยุกต์ใช้สำหรับการลงทุนในตลาดหุ้น ซึ่งได้รวบรวมมาทั้งหมดจำนวน 2,326 งานที่เกี่ยวข้อง ตั้งแต่ปี 1995 จนถึงปี 2019 จากเว็บไซต์ Scopus โดยผู้นำเสนอได้แบ่งลักษณะงานที่เกี่ยวข้องทั้งหมดออกเป็น 4 หมวดหมู่ ได้แก่ Portfolio optimization, Stock market prediction using AI, Financial sentiment Analysis และ Combination approaches สำหรับหมวดหมู่แรก คือ Portfolio optimization จะเป็นงานที่เกี่ยวข้องในการบริหารจัดการพอร์ตสำหรับนักลงทุนเพื่อให้ได้ผลกำไรที่สูงสุด (Maximum Profits) โดยเมทริกที่ใช้ในการวัดประสิทธิภาพ คือ ผลตอบแทนที่คาดหวัง (Expected return) และความเสี่ยงในผลตอบแทนที่ไม่แน่นอน (Risk) สำหรับหมวดหมู่ที่สอง คือ Stock market prediction using AI จะเป็นงานที่เกี่ยวข้องในการนำชุดข้อมูลในอดีต (Historical time series) มาสร้างโมเดลสำหรับการทำนายผ่านอัลกอริทึมทางปัญญาประดิษฐ์ (AI) ที่หลากหลาย ทั้งวิธีการเรียนรู้ของเครื่อง และวิธีการเรียนรู้เชิงลึก เป็นต้น โดยตัวแปร (Input variables) ส่วนใหญ่ที่นิยมใช้ คือ ตัวชี้วัดทางเทคนิค (Technical Indicators) ซึ่งงานที่เกี่ยวข้องนั้นได้มีทิศทางในการสำรวจและใช้เทคนิควิธีที่ซับซ้อนมากขึ้นในส่วนของขั้นตอนการเตรียมข้อมูล ซึ่งทำให้ค่าความแม่นยำสูงขึ้น เนื่องจากได้ลดขนาดข้อมูลและกำจัดข้อมูลรบกวน (Noisy data) ออกไป สำหรับหมวดหมู่ที่สาม คือ Financial sentiment Analysis จะเป็นงานที่เกี่ยวข้องในการทำนายผลผ่านวิเคราะห์อารมณ์และความรู้สึกของมวลชนในการวัดว่าความคิดเห็นหรือข้อความ (Text) นั้นเป็นไปในเชิงบวก หรือเชิงลบ (Negative or Positive) ซึ่งงานที่เกี่ยวข้องส่วนใหญ่จะดึงข้อมูลประเภทข้อความ (Texts) จากสื่อสังคมออนไลน์และข่าวจากเว็บไซต์ก่อนจะเข้าสู่วิธีการประมวลผลภาษาธรรมชาติ (NLP) สำหรับหมวดหมู่สุดท้าย คือ Combination approaches จะเป็นงานที่เกี่ยวข้องที่มีการประยุกต์ใช้เทคนิควิธีมากกว่าหนึ่งวิธีขึ้นไป เช่น การใช้วิธีการเรียนรู้ของเครื่องทำนายผลร่วมกับการวิเคราะห์โดยใช้วิธี Sentiment analysis เป็นต้น จากที่กล่าวมาข้างต้น ผู้เสนอได้ข้อสรุปว่า งานวิจัยและงานที่เกี่ยวข้องในการใช้ปัญญาประดิษฐ์สำหรับประยุกต์ใช้ในการลงทุนทางการเงิน ได้มีจำนวนชิ้นงานและการศึกษาที่เพิ่มขึ้นอย่างต่อเนื่องแบบก้าวกระโดด (Exponential increase) ตั้งแต่ปี 1995 จนถึงปี 2019 ซึ่งสะท้อนให้

เห็นว่าการศึกษานี้ได้รับความสนใจเพิ่มมากขึ้นในปัจจุบัน โดยหมวดหมู่ Combination approaches หรือการประยุกต์ใช้เทคนิควิธีมากกว่าหนึ่งเทคนิคขึ้นไป แสดงผลได้ค่อนข้างเหนือกว่าการใช้เทคนิควิธีเดียว ซึ่งจำนวนงานวิจัยที่เกี่ยวข้องในหมวดหมู่นี้ยังคงมีจำนวนไม่มากนัก [19]

Pattern graph tracking-based stock price prediction using big data (2018)

Seungwoo Jeon, Bonghee Hong และ Victor Chang ได้เสนอวิธีการแบบใหม่ (Novel method) ในการทำนายราคาหุ้น โดยผู้เสนอได้ใช้อัลกอริทึม Dynamic Time Warping (DTW) เพื่อค้นหารูปแบบราคาหุ้นในอดีตที่คล้ายคลึงกับรูปแบบในปัจจุบัน ควบคู่กับ Stepwise Regression Analysis ในการคัดเลือกฟีเจอร์ (Feature selection) เพื่อหาตัวแปรที่มีผลกระทบต่อราคาหุ้นมากที่สุด และใช้โมเดลโครงข่ายประสาทเทียม (Artificial Neural Network) โดยใช้ตัวแปรที่ผ่านขั้นตอนการคัดเลือกฟีเจอร์ เป็นตัวแปรสำหรับฝึกสอนในโมเดลสำหรับการทำนายราคาหุ้นแบบรายวัน (Daily) [20]

Improving S&P stock prediction with time series stock similarity (2020)

Lior Sidi ได้นำเสนอเทคนิคการทำนายแนวโน้มดัชนีราคาหุ้น S&P500 (ข้อมูลย้อนหลังตั้งแต่ปี 2012 ถึงปี 2017) โดยใช้เทคนิคการหาความคล้ายคลึงกันของอนุกรมเวลาเพิ่มเข้ามาในระหว่างการฝึกสอนโมเดลและการทำนายผล ด้วยสมมุติฐานที่ว่าวิธีการที่นำเสนอจะสามารถพัฒนาประสิทธิภาพของโมเดลในการทำนายได้ โดยตัวแปรที่ผู้เสนอใช้ในการทดลองคือดัชนีชี้วัดทางเทคนิค ได้แก่ MACD, RSI, PROC, Volume และอื่นๆ เป็นต้น รวมถึงใช้เทคนิคการลดมิติและปรับขนาดข้อมูลในขั้นตอนของการเตรียมข้อมูล นอกจากนี้ยังได้ทดลองศึกษาโมเดลแบบหนึ่งตัวแปร (Univariate) และโมเดลแบบหลายตัวแปร (Multivariate) ในการสร้างโมเดลสำหรับทำนายผล โดยโมเดลที่ผู้เสนอใช้ทดลองศึกษาคือ Ensemble algorithms ได้แก่ Random Forest และ Gradient Boosting Trees ซึ่งเป้าหมายในการทำนายผลของการทดลองนี้ คือการทำนายว่าราคาหุ้นจะเพิ่มขึ้นหรือลดลงในช่วงเวลาถัดไป จากผลการทดลอง ผู้เสนอได้ข้อสรุปว่า โมเดลที่ได้มีการพัฒนาด้วยเทคนิควิธีที่เสนอมีประสิทธิภาพที่ดีกว่าโมเดลที่ใช้วิธีการอื่นๆ ในส่วนของค่าความแม่นยำ (Accuracy) และผลกำไร (Profit) โดยมีค่าเฉลี่ยของค่าความแม่นยำอยู่ที่ 55% ซึ่งมีค่าที่สูงกว่าเมื่อเทียบกับโมเดลอื่นๆ และมีผลกำไรเฉลี่ยอยู่ที่ 19.87 ซึ่งมีค่าที่สูงกว่าเมื่อเทียบกับโมเดลอื่นๆเช่นเดียวกัน [21]

Stock Market Trend Prediction Using High-Order Information of Time Series (2019)

Min Wen, Ping Li, Lingfei Zhang และ Yan Chen ได้นำเสนออัลกอริทึม Motif-based sequence reconstruction ในการดึงรูปแบบของราคาหุ้นที่คล้ายกัน ควบคู่กับโมเดล Convolutional Neural Network ในการทำนายแนวโน้มของราคาหุ้นในวันถัดไป (Daily) บนชุดข้อมูลตลาดหุ้นจาก Yahoo Finance Website โดยตัวแปรหลักที่ใช้ในการฝึกสอนโมเดลคือ ราคาปิด (Close) จากผลการทดลอง ผู้เสนอได้ข้อสรุปว่า ประสิทธิภาพของวิธีการที่ได้นำเสนอนั้นมีค่าความแม่นยำเพิ่มสูงขึ้นประมาณ 4-7% เมื่อเปรียบเทียบกับประสิทธิภาพของวิธีการแบบดั้งเดิม ได้แก่ ARIMA และ Wavelet วิธีการเรียนรู้ของเครื่อง ได้แก่ Hidden Markov Model (HMM) และวิธีการเรียนรู้เชิงลึก ได้แก่ Long Short-Term Memory (LSTM) และ Structure From Motion (SFM) ในการทำนายแนวโน้มราคาหุ้น [22]

Stock Trend Prediction with Technical Indicator using SVM (2014)

Xinjie Di ได้นำเสนอกรอบการทำนายแนวโน้มราคาหุ้น AAPL, MSFT และ AMZN ในช่วงระยะสั้น (ช่วงแนวโน้ม 1-10 วัน) โดยใช้ตัวชี้วัดทางเทคนิค (Technical Indicators) เป็นตัวแปรนำเข้า และได้ใช้ Tree algorithm ในการคัดเลือกฟีเจอร์ (Feature selection) เพื่อหาตัวแปรอินดิเคเตอร์ที่สำคัญที่มีผลต่อการทำนายแนวโน้มราคาหุ้น ซึ่งคลาสเป้าหมายนั้นประกอบด้วย แนวโน้มขาขึ้น และ แนวโน้มขาลง โดยในส่วนของโมเดลทำนายผล ผู้เสนอได้ใช้ RBF-Kernelized SVM Model และการทำ Grid Search เพื่อปรับจูนไฮเพอร์พารามิเตอร์ของโมเดล ผู้เสนอได้ข้อสรุปว่าวิธีการที่ผู้เสนอได้ทดลองสามารถแสดงค่า Accuracy เฉลี่ยได้สูงกว่า 70% สำหรับการทำนายแนวโน้มในช่วง 3-10 วัน ผ่านอัลกอริทึม RBF-Kernelized SVM โดยสำหรับการทำนายแนวโน้มในช่วงวันถัดไป (1 วัน) โมเดลได้แสดงค่าความแม่นยำเฉลี่ยอยู่ที่ 56% ซึ่งก็ถือว่ามีประสิทธิภาพเช่นเดียวกัน เนื่องจากการทำนายแนวโน้มในวันถัดไปนั้นค่อนข้างที่จะมีแนวโน้มการเปลี่ยนแปลงแบบสุ่ม (Random) ซึ่งวิธีการของผู้เสนอสามารถเอาชนะในส่วนนี้ได้ [23]

Stock Market Prediction with Gaussian Naive Bayes Machine Learning Algorithm (2021)

Ernest Kwame Ampomah, Gabriel Nyame, Zhiguang Qin, Prince Clement Addo, Enoch Opanin Gyamfi และ Micheal Gyan ได้นำเสนออัลกอริทึมการเรียนรู้ของเครื่องการจำแนกแบบเบย์อย่างง่าย (Gaussian Naive Bayes) ควบคู่กับเทคนิคการทำ Feature scaling และ Feature extraction ที่แตกต่างกัน บนชุดข้อมูลตลาดหุ้นจาก NYSE, NASDAQ และ NSE ในการทำนายแนวโน้มราคาหุ้นในวันถัดไป โดยตัวแปรหลักที่ใช้ในการฝึกสอนโมเดลคือ ดัชนีชี้วัดทางเทคนิค (Technical Indicators) ทั้งหมด 40 ตัว เพื่อประเมินผลและเปรียบเทียบเทคนิคที่ให้ประสิทธิภาพสูงสุด จากผลการทดลอง ผู้เสนอได้ข้อสรุปว่า แบบจำลอง Gaussian Naive Bayes ควบคู่กับเทคนิคการปรับขนาดคุณลักษณะแบบ MinMax Normalization และเทคนิคการสกัดคุณลักษณะแบบ Linear Discriminant Analysis (LDA) แสดงประสิทธิภาพได้ดีที่สุดเมื่อเทียบกับวิธีการอื่นๆ [24]

Short and Long Term Stock Trend Prediction using Decision Tree (2017)

Rupesh A. Kamble ได้นำเสนอวิธีการทดลองในการทำนายแนวโน้มราคาหุ้นจาก NSE และ BSE เป็นเวลาย้อนหลัง 5 ปี โดยสำหรับในขั้นตอนการเตรียมข้อมูลและการเทรนโมเดลจะใช้ข้อมูล 3 ปี ย้อนหลัง โดยใช้ราคาปิด (Closing price) เป็นคลาสเป้าหมายทั้งในส่วนของขั้นตอนการเทรนโมเดล และทดสอบโมเดล ซึ่งในจุดประสงค์แรกของการศึกษาจะเป็นการพัฒนาวิธีการทำนายแนวโน้มราคาหุ้นในระยะสั้นให้มีประสิทธิภาพมากยิ่งขึ้นโดยใช้อินดิเคเตอร์ ได้แก่ MACD, RSI, KDJ และ BB ซึ่งจากการทดลองสามารถสังเกตได้ว่า การเลือกใช้เทคนิควิธีในการเตรียมข้อมูล และการเลือกใช้โมเดลการเรียนรู้ของเครื่องที่เหมาะสม จะสามารถเพิ่มอัตราค่า Accuracy สำหรับการทำนายแนวโน้มราคาหุ้นในระยะสั้นได้ สำหรับการทำนายแนวโน้มราคาหุ้นในระยะยาว การใช้อินดิเคเตอร์เพียงอย่างเดียวไม่น่าเพียงพอ แต่ผลจากการทดลองพบว่า การใช้ตัวแปรทั้งในทางเทคนิคและทางพื้นฐานของตัวหุ้นนั้น มีความเป็นไปได้ที่จะสามารถทำนายแนวโน้มราคาหุ้นในระยะยาวได้ [25]

2.3 เทคโนโลยีที่เกี่ยวข้อง

2.3.1 ภาษาโปรแกรมไพทอน (Python Programming Language)

ภาษาโปรแกรมไพทอน (Python) คือ ภาษาโปรแกรมคอมพิวเตอร์ระดับสูงที่ถูกออกแบบมาให้เป็นภาษาสคริปต์ที่อ่านง่าย โดยตัดความซับซ้อนของโครงสร้างและไวยากรณ์ของภาษาออกไป ในส่วนของการแปลงชุดคำสั่งที่เราเขียนให้เป็นภาษาเครื่อง ไพทอนมีการทำงานแบบ Interpreter คือเป็นการแปลชุดคำสั่งทีละบรรทัด เพื่อป้อนเข้าสู่หน่วยประมวลผลให้คอมพิวเตอร์ทำงานตามที่เราต้องการ นอกจากนั้นภาษาโปรแกรม Python ยังสามารถนำไปใช้ในการเขียนโปรแกรมได้หลากหลายประเภท โดยไม่ได้จำกัดอยู่ในงานเฉพาะทางใดทางหนึ่ง (General-purpose language) จึงทำให้มีการนำไปใช้กันแพร่หลายในหลายองค์กรใหญ่ระดับโลก เช่น Google, YouTube, Instagram, Dropbox และ NASA เป็นต้น [26]

2.3.2 ไลบรารีที่เกี่ยวข้อง (Library)

ไลบรารี (Library) เปรียบเสมือนโปรแกรมสำเร็จรูปที่เก็บฟังก์ชันการทำงานที่เฉพาะทางประโยชน์ของไลบรารีสำเร็จรูปที่มากับโปรแกรมก็คือผู้ใช้งานไม่จำเป็นต้องสร้างไลบรารีขึ้นมาใหม่เองทั้งหมด แต่สามารถนำไลบรารีที่ได้มีการพัฒนาไว้อยู่แล้วมาใช้งานได้เลย

2.3.2.1 numpy

numpy เป็นแพ็คเกจพื้นฐานสำหรับการคำนวณเชิงวิทยาศาสตร์ผ่านภาษาโปรแกรมไพทอน ซึ่งเป็นไลบรารีที่มีการจัดเตรียมวัตถุอาร์เรย์หลายมิติ วัตถุที่ได้รับประเภทต่างๆ เช่น อาร์เรย์แบบปัดและเมทริกซ์ และชุดของรูทีนที่มีการดำเนินการอย่างรวดเร็ว ได้แก่ คณิตศาสตร์ ตรรกศาสตร์ การจัดการรูปทรง การเรียงลำดับ การเลือก I/O การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง พีชคณิตเชิงเส้นพื้นฐาน การดำเนินการทางสถิติพื้นฐาน การจำลองแบบสุ่ม และอื่นๆอีกมากมาย เป็นต้น [27]

2.3.2.2 pandas

pandas เป็นแพ็คเกจสำหรับภาษาโปรแกรมไพทอนที่มีโครงสร้างข้อมูลที่รวดเร็ว ยืดหยุ่น ซึ่งถูกออกแบบมาเพื่อให้สามารถทำงานได้กับข้อมูลที่เป็นแบบ Relational หรือ

Labeled ได้ง่าย โดยไลบรารีนี้ถูกสร้างมาเพื่อเป็นโครงสร้างระดับสูงขั้นพื้นฐานสำหรับการวิเคราะห์ข้อมูล ทั้งการวิเคราะห์ในเชิงปฏิบัติ (Practical) และวิเคราะห์บนโลกเป็นความจริง (Real-time) ผ่านภาษาโปรแกรมไพทอน นอกจากนี้ยังมีเป้าหมายที่กว้างขวางขึ้นในการเป็นเครื่องมือสำหรับการวิเคราะห์และจัดการข้อมูล Open source ที่มีประสิทธิภาพผ่านภาษาโปรแกรมใดก็ได้ [28]

2.3.2.3 matplotlib

Matplotlib เป็นไลบรารีข้ามแพลตฟอร์ม (cross-platform) ในการพล็อตกราฟ และการแสดงผลข้อมูลสำหรับภาษาโปรแกรมไพทอนและไลบรารี Numpy ซึ่งเป็นทางเลือกแบบ Open source ในการใช้งานแทนที่ Matlab โดยนักพัฒนาทั่วไปสามารถใช้ API ของ Matplotlib ในการฝังพล็อตใน GUI Applications ได้ [29]

2.3.2.4 seaborn

Seaborn เป็นไลบรารีในภาษาไพทอนที่ใช้สำหรับการแสดงผลข้อมูล (Data Visualization) บน Matplotlib ซึ่งมีอินเทอร์เฟซระดับสูงสำหรับการวาดกราฟฟิกทางสถิติที่น่าสนใจและเป็นประโยชน์ในการวิเคราะห์ผล [30]

2.3.2.5 sklearn

Scikit-learn หรือ sklearn เป็นไลบรารีที่มีประโยชน์สำหรับ Machine Learning ในภาษาโปรแกรมไพทอน ซึ่งมีเครื่องมือที่มีประสิทธิภาพให้เลือกใช้งานสำหรับวิธีการเรียนรู้ของเครื่องและการสร้างโมเดลทางสถิติ ได้แก่ Classification, Regression, Clustering และ Dimensionality Reduction ผ่านอินเทอร์เฟซในภาษาไพทอน [31]

2.3.2.6 yfinance

yfinance เป็นไลบรารีที่สามารถเรียกใช้งานได้ผ่านภาษาโปรแกรมไพทอนจาก Yahoo Finance ซึ่งเป็นแหล่งที่รวบรวมและนำเสนอข้อมูลทางการตลาดที่หลากหลาย ได้แก่ หลักทรัพย์ พันธบัตร สกุลเงิน และคริปโตเคอร์เรนซี [32]

2.3.2.7 dtaidistance

dtaidistance เป็น Time Series Distance Library ซึ่งเป็นไลบรารีสำหรับเรียกใช้งานเพื่อคำนวณระยะห่างระหว่างข้อมูลอนุกรมเวลา เช่น Dynamic Time Warping ซึ่งได้ถูกใช้งานใน DTAI Research Group [33]

2.3.3 เครื่องมือ (Tools)

2.3.3.1 Jupyter Notebook

Jupyter Notebook คือ เครื่องมือหนึ่งที่นิยมมากในด้านของ Data Science ซึ่งงานด้านนี้ต้องทำงานที่เกี่ยวกับการจัดการข้อมูลเป็นจำนวนมากๆ แล้วก็ต้องรายงานงานวิจัยที่วิจัยไว้ ซึ่งตัว Jupyter Notebook ก็ได้ออกแบบมาตรงตามจุดประสงค์ของการใช้งานไม่ว่าจะเป็น การเรียกใช้งาน library พร้อมทั้งเขียน code และดูผลได้เลย Jupyter Notebook นั้นถูกออกแบบมาให้ทำงานและอ่านได้ง่ายกว่าการใช้งานโปรแกรมแบบปกติ [34]

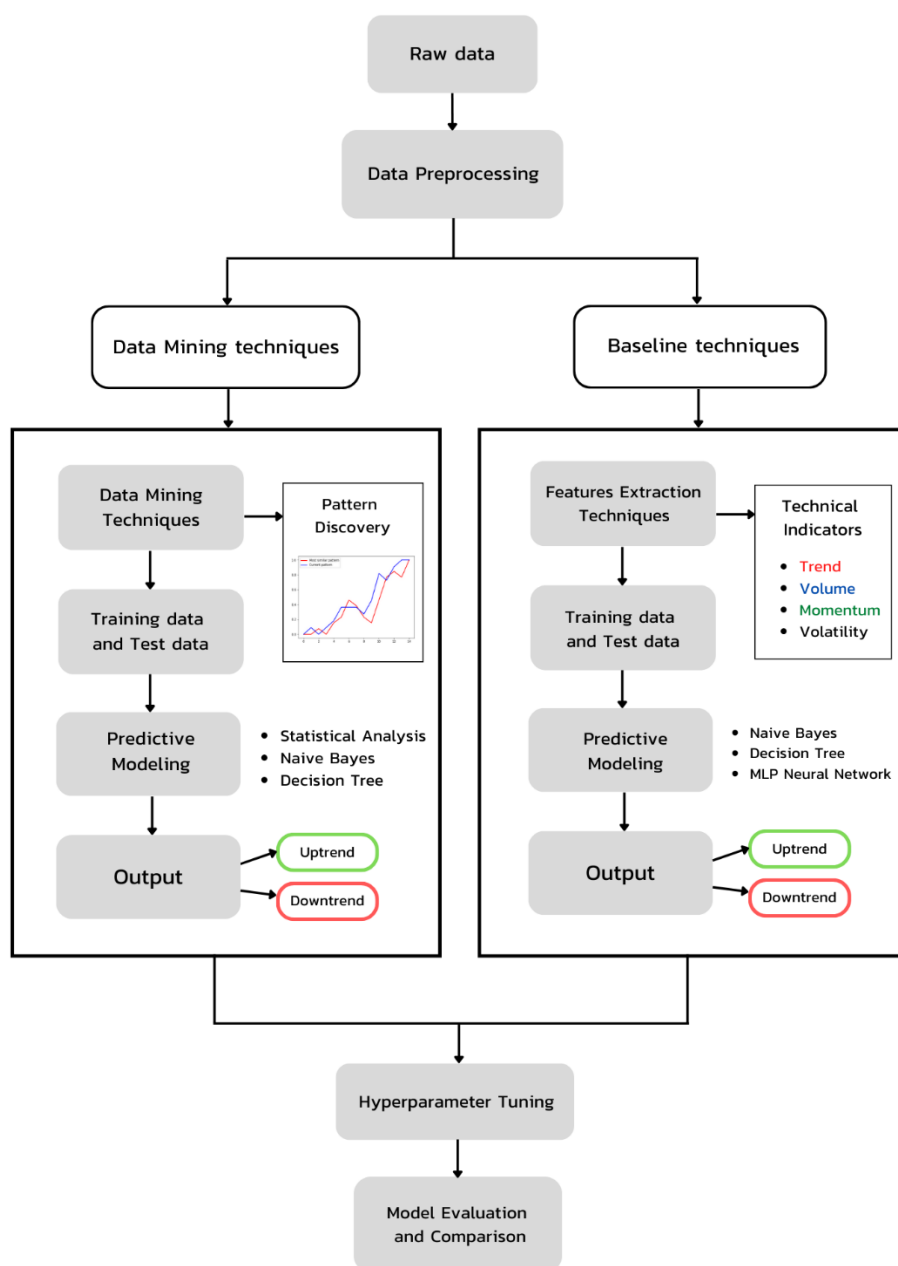
2.3.3.1 Google Colab

Google Colab คือ Jupyter notebook ดัดแปลงที่รันอยู่บนคลาวด์ และไม่จำเป็นต้องติดตั้งโปรแกรมใดๆ ก่อนใช้งาน ซึ่งทุกคนสามารถใช้งานได้ เพียงแค่มีบัญชี Google Drive เพื่อใช้ในการจัดเก็บตัวโค้ดเท่านั้น โดยภาษา Python เป็นภาษาหลักที่ใช้ในการเขียนและรันงานบนเว็บ Colab นี้ [35]

บทที่ 3

วิธีการวิจัย

3.1 ภาพรวมของการศึกษา



ภาพที่ 3.1 แผนภาพแสดงภาพรวมของการศึกษา

จากภาพรวมการศึกษาแสดงให้เห็นถึงขั้นตอนการดำเนินงานโดยเริ่มต้นตั้งแต่ขั้นตอนการเก็บรวบรวมข้อมูลดิบในแต่ละรายการ ขั้นตอนการจัดเตรียมข้อมูล รวมถึงขั้นตอนการทำเหมืองข้อมูล เพื่อสร้างชุดข้อมูลสำหรับฝึกสอนให้โมเดลเรียนรู้ จากนั้นทำการทดสอบโมเดลเพื่อประเมินประสิทธิภาพของความแม่นยำในการทำนายผล ขั้นตอนสุดท้ายทำการเปรียบเทียบเทคนิควิธีการทำเหมืองข้อมูลที่ได้ทดลองศึกษากับเทคนิควิธีการพื้นฐานบนชุดข้อมูลเดียวกัน ผ่านขั้นตอนการดำเนินงานในลักษณะเดียวกัน

3.2 วิธีการศึกษา

3.2.1 ขั้นตอนการรวบรวมข้อมูล (Data Collection)

ชุดข้อมูลตลาดหุ้นที่ใช้การศึกษา คือ ชุดข้อมูลตลาดหุ้นจากตลาดหลักทรัพย์แห่งประเทศไทย (The Stock Exchange of Thailand : SET) ในกลุ่มอุตสาหกรรมที่แตกต่างกันทั้งหมด 5 รายการ ได้แก่ KBANK (กลุ่มธุรกิจการเงิน), PTT (กลุ่มทรัพยากร), CPALL (กลุ่มบริการ), DELTA (กลุ่มเทคโนโลยี) และ SCC (กลุ่มอสังหาริมทรัพย์และก่อสร้าง) โดยในส่วนขั้นตอนวิธีการรวบรวมชุดข้อมูล สามารถดึงชุดข้อมูลได้โดยเรียกผ่านไลบรารี Yahoo Finance ผ่านภาษาโปรแกรมไพทอน โดยชุดข้อมูลทั้งหมดของแต่ละรายการจะแสดงตัวแปรเริ่มต้นพร้อมข้อมูลในแต่ละคอลัมน์ทั้งหมด 7 คอลัมน์ ได้แก่ 1. เวลา (Date) 2. ราคาเปิด (Open) 3. ราคาสูงสุด (High) 4. ราคาต่ำสุด (Low) 5. ราคาปิด (Close) 6. ราคาปิดที่ถูกปรับแล้ว (Adj Close) และ 7. ปริมาณซื้อขาย (Volume)

	Open	High	Low	Close	Adj Close	Volume
Date						
2018-10-11 00:00:00+07:00	51.25	52.00	50.75	51.0	42.419773	161422900
2018-10-12 00:00:00+07:00	51.50	52.00	51.00	51.5	42.835648	78177000
2018-10-16 00:00:00+07:00	51.50	51.75	51.00	51.0	42.419773	43143300
2018-10-17 00:00:00+07:00	51.50	51.75	51.00	51.0	42.419773	39322500
2018-10-18 00:00:00+07:00	50.75	51.50	50.50	51.0	42.419773	42968200

ภาพที่ 3.2 ตัวอย่างข้อมูลดิบของรายการหลักทรัพย์ (Raw data)

3.2.2 ขั้นตอนการเตรียมข้อมูล (Data Preprocessing)

3.2.2.1 การทำข้อมูลให้เป็นบรรทัดฐาน (Data Normalization)

เป็นขั้นตอนในการปรับข้อมูลให้เป็นบรรทัดฐาน (Normalization) สำหรับชุดข้อมูลหลักทรัพย์ที่ได้ทำการเก็บรวบรวม เนื่องจากข้อมูลหลักทรัพย์นั้นมีความผันผวนของราคาค่อนข้างสูงและมีลักษณะไม่คงที่ ซึ่งมีช่วงของราคาที่ไม่แน่นอน และสามารถเปลี่ยนแปลงได้อย่างมีนัยยะสำคัญ ส่งผลให้มีสัดส่วน (Scale) ของข้อมูลในแต่ละคอลัมน์ต่างกัน จึงต้องทำการปรับขนาดข้อมูลให้อยู่ในช่วงเดียวกัน โดยในการศึกษานี้จะทำการปรับขนาดข้อมูลด้วย MinMax Normalization (ปรับให้ช่วงของข้อมูลมีค่าอยู่ระหว่าง 0-1) เนื่องจากเป็นวิธีในการปรับขนาดข้อมูลที่มีประสิทธิภาพบนชุดข้อมูลหลักทรัพย์อ้างอิงจากงานที่เกี่ยวข้อง

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

โดยแทน x คือ ราคา (Close) ในคอลัมน์

$\min(x)$ คือ ราคาที่สูงที่สุด (Minimum Close) ในคอลัมน์

$\max(x)$ คือ ราคาที่ต่ำที่สุด (Maximum Close) ในคอลัมน์

3.2.2.2 การแปลงข้อมูล (Data Transformation)

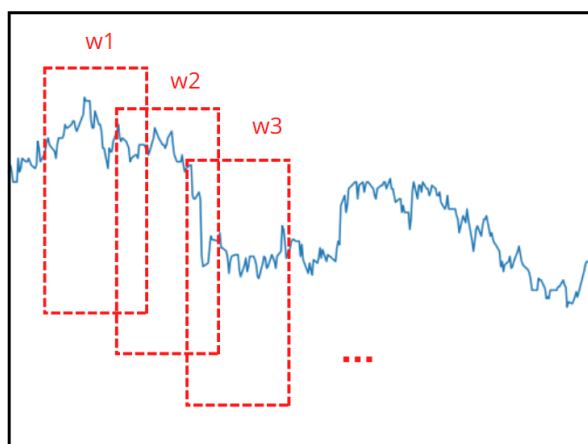
เป็นขั้นตอนในการแปลงข้อมูล โดยได้ทำการแปลงคอลัมน์เวลา (Date) ซึ่งอยู่ในรูปแบบ Index ในตอนแรกให้เป็นรูปแบบเดียวกับคอลัมน์อื่นๆ และสร้างคอลัมน์ Movement ของราคาซึ่งมีข้อมูลประกอบด้วย '+1' และ '-1' โดยแปลงมาจากการเทียบข้อมูลราคาปิดในคอลัมน์ Close ในแถวก่อนหน้าเทียบกับแถวปัจจุบัน จะได้เป็นคอลัมน์ Target class ซึ่งเก็บข้อมูลแนวโน้มราคาในวันถัดไป

3.2.3 ขั้นตอนการทำเหมืองข้อมูล (Data Mining)

เป็นขั้นตอนในการดึงคุณลักษณะหรือตัวแปรที่สำคัญผ่านการทำเหมืองข้อมูลเพื่อค้นหารูปแบบของแนวโน้มในชุดข้อมูลหลักทรัพยากรย้อนหลังที่มีลักษณะคล้ายคลึงกับรูปแบบแนวโน้มในปัจจุบัน ซึ่งในการศึกษานี้จะทำการดึงคุณลักษณะทั้งหมดเพื่อกำหนดเป็นตัวแปรในการใช้ทำนายแนวโน้มของราคาในช่วงวันถัดไป โดยมีอัลกอริทึมและเทคนิควิธีที่เกี่ยวข้องดังต่อไปนี้

3.2.3.1 Sliding Windows Technique

เป็นวิธีการแบ่งช่วงข้อมูลอนุกรมเวลาในแต่ละลำดับย่อย (Subsequence) ทั้งหมดของชุดข้อมูลในอดีตของราคาหุ้น โดยมีการกำหนดขนาด (Window size) ของลำดับย่อยตามค่าที่ได้กำหนดไว้

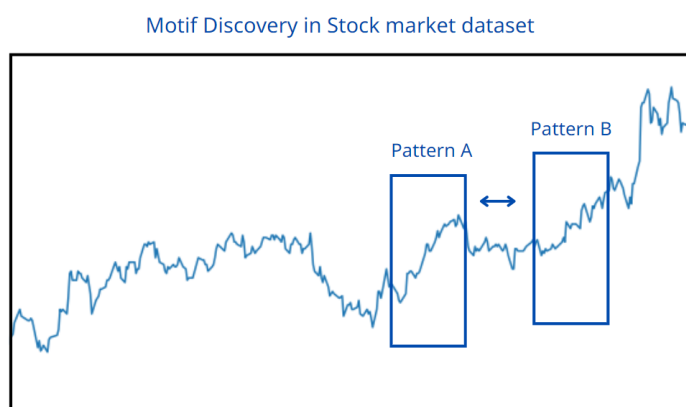


ภาพที่ 3.3 ตัวอย่าง Sliding window technique บนข้อมูลอนุกรมเวลา

3.2.3.2 การค้นพบโมทีฟ (Motif Discovery)

เป็นเทคนิควิธีการที่ใช้เพื่อค้นหารูปแบบแนวโน้มราคา (Price Patterns) ในชุดข้อมูลในอดีตที่มีลักษณะคล้ายคลึงกับแนวโน้มรูปแบบราคาในปัจจุบัน ผ่านกระบวนการคำนวณหาค่าระยะห่างความคล้ายกันของลำดับย่อยอนุกรมเวลา (Similarity Distance Measure) โดยใช้ อัลกอริทึม Dynamic Time Warping

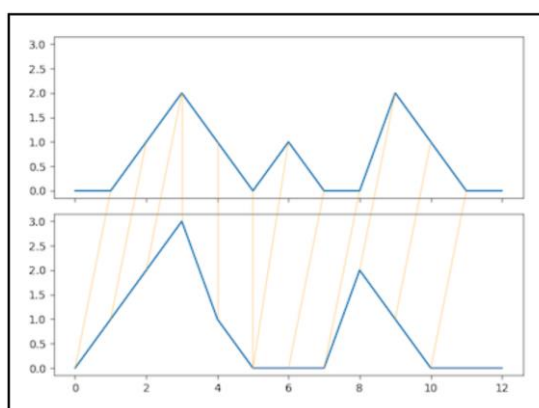
เนื่องจากเป็นอัลกอริทึมที่มีประสิทธิภาพดีกว่าอัลกอริทึมอื่นๆ ในการคำนวณบนข้อมูลราคาหลักทรัพย์อ้างอิงจากงานที่เกี่ยวข้อง



ภาพที่ 3.4 ตัวอย่างวิธีการค้นพบโมทีฟบนข้อมูลอนุกรมเวลา

3.2.3.3 Dynamic Time Warping Algorithm

เป็นอัลกอริทึมที่ใช้ในการคำนวณระหว่างความคล้ายคลึงกันระหว่างรูปแบบราคาในอดีตที่มีความคล้ายคลึงกับรูปแบบปัจจุบัน



ภาพที่ 3.3 ตัวอย่างการ Matching ของ DTW algorithm ที่ใช้คำนวณหาความคล้ายกันระหว่างรูปแบบราคา

3.2.4 ขั้นตอนการสร้างวิธีการทำนายผล (Predictive Modeling)

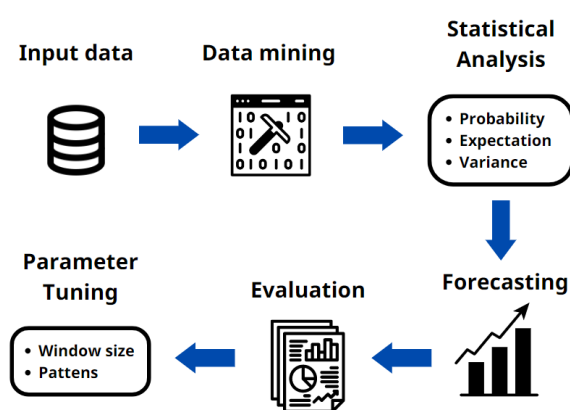
3.2.4.1 การแบ่งชุดข้อมูลฝึกสอนและทดสอบ (Training and Testing data)

เป็นขั้นตอนในการแบ่งชุดข้อมูลก่อนเข้าสู่ขั้นตอนการวิเคราะห์และประมวลผลผ่านโมเดลการทำนายผล โดยในการศึกษานี้จะทำการแบ่งชุดข้อมูลโดยเป็นชุดข้อมูลสำหรับฝึกสอน 85% และชุดข้อมูลสำหรับทดสอบ 15% ซึ่งจะเป็นการแบ่งชุดข้อมูลเรียงตามลำดับเวลาแบบอนุกรมเวลา (Time series) คือ ชุดข้อมูลฝึกสอนคือรูปแบบของราคาในอดีต และชุดข้อมูลทดสอบคือรูปแบบราคาในปัจจุบัน

3.2.4.2 การสร้างโมเดลวิธีการทำนายผล (Predictive Modeling)

สำหรับในขั้นตอนนี้จะเป็นการนำวิธีการและโมเดลการทำนายผลมาสร้างกรอบวิธีสำหรับการทำนายแนวโน้มของราคา ซึ่งประกอบด้วย 2 คลาส ได้แก่ แนวโน้มราคาขึ้น (Uptrend) และแนวโน้มราคาลง (Downtrend) โดยวิธีการทำนายผลที่จะทำการศึกษา มีดังนี้

(1) การวิเคราะห์ทางสถิติ (Statistical Analysis)



ภาพที่ 3.5 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีการวิเคราะห์ทางสถิติ

เป็นวิธีในการวิเคราะห์พฤติกรรมของรูปแบบการเคลื่อนไหวของราคาในอดีตที่มีลักษณะคล้ายกับรูปแบบการเคลื่อนไหวของราคาปัจจุบัน

ซึ่งมีการเกิดขึ้นอย่างซ้ำๆต่อเนื่อง โดยสามารถนำรูปแบบการเคลื่อนไหวของราคาที่ได้ มาทำการวิเคราะห์และคาดการณ์แนวโน้มผ่านวิธีการวิเคราะห์ทางสถิติได้ดังนี้

3.2.4.2.1.1 การหาค่าความน่าจะเป็นของเหตุการณ์

เป็นวิธีการพื้นฐานที่ง่ายที่สุดในการหาแนวโน้มการเปลี่ยนแปลงของราคาในอนาคต โดยสามารถวิเคราะห์ได้จากการหาจำนวนแนวโน้มการเปลี่ยนแปลงของราคาที่เราสนใจในอดีตส่วนด้วยจำนวนแนวโน้มการเปลี่ยนแปลงของราคาทั้งหมด จากจำนวนรูปแบบการเคลื่อนไหวของราคาในอดีตที่ได้

$$\text{ความน่าจะเป็นของแนวโน้มราคาในวันถัดไป} = \frac{\text{จำนวนแนวโน้มราคาในอดีตที่สนใจ}}{\text{จำนวนแนวโน้มราคาในอดีตทั้งหมด}}$$

3.2.4.2.1.2 การหาค่าความคาดหวังของตัวแปรสุ่ม

สามารถวิเคราะห์ได้จากการกำหนดตัวแปรสุ่มของแนวโน้มราคา คือ แนวโน้มราคาขึ้น (+1) และแนวโน้มราคาลง (-1) เพื่อหาค่าเฉลี่ยโอกาสของแนวโน้มราคาที่จะเกิดขึ้นในวันถัดไป ซึ่งคำนวณได้จาก

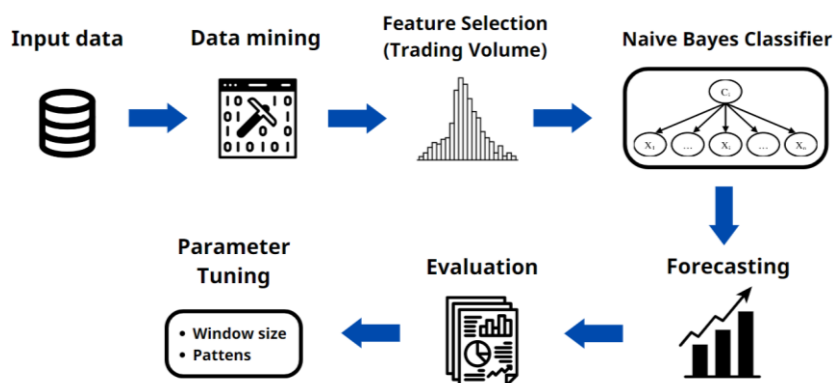
$$\begin{aligned} \text{ค่าความคาดหวังของแนวโน้มราคาในวันถัดไป} = \\ \sum (\text{ตัวแปรสุ่มของแนวโน้มราคา} \times \text{ความน่าจะเป็นที่จะเกิดแนวโน้มราคา}) \end{aligned}$$

3.2.4.2.1.3 การหาค่าความแปรปรวนของตัวแปรสุ่ม

สามารถวิเคราะห์โดยคำนวณหาค่าความแปรปรวนของตัวแปรสุ่มแนวโน้มราคา เพื่อใช้วัดการกระจายตัวของการเคลื่อนที่ราคาของรูปแบบในอดีตรอบค่าเฉลี่ยของแนวโน้มราคา ซึ่งสามารถนำมาวิเคราะห์ความเสี่ยงและความผันผวนของแนวโน้มราคาที่เกิดขึ้นได้ ซึ่งคำนวณได้จาก

$$\begin{aligned} \text{ค่าความแปรปรวนของแนวโน้มราคาในวันถัดไป} = \\ E[(\text{ตัวแปรสุ่มของแนวโน้มราคา} - \text{ค่าความคาดหวังของแนวโน้มราคาในวันถัดไป})^2] \end{aligned}$$

(2) วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 3.6 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีจำแนกแบบเบย์อย่างง่าย

เป็นวิธีการทำนายแบบ Classification ในการนำมาใช้ทำนายแนวโน้มราคาในวันถัดไปโดยใช้หลักความน่าจะเป็นเข้ามาช่วยในการคำนวณ โดยเมื่อผ่านกระบวนการทำเหมืองข้อมูลบนราคาหลักทรัพย์แล้วจะเข้าสู่ขั้นตอนการคัดเลือกฟีเจอร์เพื่อเลือกตัวแปรในการสร้างโมเดลทำนายผล ซึ่งก็คือปริมาณซื้อขายรายวันจากรูปแบบของราคาหลักทรัพย์ทั้งหมด โดยเมื่อเตรียมข้อมูลในขั้นตอนนี้เสร็จแล้วจะเข้าสู่ขั้นตอนถัดไปคือการสร้างโมเดลวิธีจำแนกแบบเบย์อย่างง่าย ดังนี้

$$P(\text{แนวโน้มราคาในวันถัดไป} \mid \text{ปริมาณซื้อขายของรูปแบบราคาปัจจุบัน}) = \frac{P(\text{ปริมาณซื้อขายของรูปแบบราคาในอดีต} \mid \text{แนวโน้มราคา}) P(\text{แนวโน้มราคาที่เกิดขึ้นทั้งหมด})}{P(\text{ปริมาณซื้อขายของรูปแบบราคาในอดีต})}$$

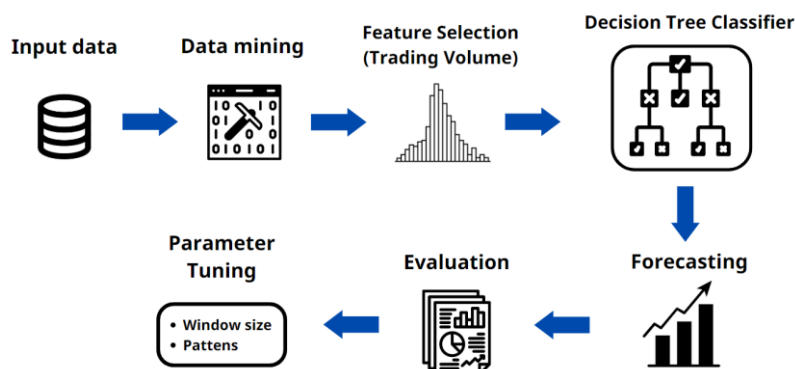
โดยที่

แนวโน้มราคาในวันถัดไป = [*Uptrend*, *Downtrend*]

ปริมาณซื้อขายของรูปแบบราคาปัจจุบัน = *Volume(Current patterns)*

ปริมาณซื้อขายของรูปแบบราคาในอดีต = *Volume(Similar patterns)*

(3) วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)



ภาพที่ 3.7 ภาพรวมของกระบวนการสร้างวิธีทำนายผลด้วยวิธีจำแนกต้นไม้การตัดสินใจ

เป็นวิธีการทำนายแนวโน้มราคาในวันถัดไปโดยใช้โมเดลต้นไม้การตัดสินใจแบบ Classifier ซึ่งเมื่อผ่านขั้นตอนการทำเหมืองข้อมูลแล้วจะเข้าสู่ขั้นตอนการคัดเลือกฟีเจอร์เพื่อเลือกตัวแปรในการสร้างโมเดลทำนายผล ซึ่งก็คือปริมาณซื้อขายรายวันจากรูปแบบของราคาหลักทรัพย์ทั้งหมด จากนั้นจะเข้าสู่ขั้นตอนการสร้างวิธีจำแนกต้นไม้การตัดสินใจซึ่งจะใช้ฟังก์ชัน Gini Index ในการแยกการปะปนกันของข้อมูลผ่านฟังก์ชัน Gain ในการเลือกฟีเจอร์ซึ่งก็คือปริมาณซื้อขายในแต่ละวันของรูปแบบราคาที่คล้ายกันในอดีต มาสร้างเป็น Node ในต้นไม้ โดยจะคัดเลือกฟีเจอร์ที่มีค่า Gain สูงสุดมาสร้างโมเดลต้นไม้การตัดสินใจ ซึ่งมีการคำนวณดังนี้

$$Gain(S, A) = Gini(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} \left(1 - \sum_{i=1}^c P_{iv}\right)$$

โดยที่ S คือ ข้อมูลทั้งหมด

A คือ ตัวแปร (Features)

$value(A)$ คือ ประเภทในตัวแปร

V คือ สมาชิกประเภทในตัวแปร

S_v คือ ข้อมูลประเภท V ที่อยู่ในตัวแปร A

3.2.5 ขั้นตอนการวัดประสิทธิภาพ (Model Evaluation)

ขั้นตอนสำหรับการวัดประสิทธิภาพของโมเดลและวิธีการทำนายผลที่ได้เสนอ ซึ่งในการศึกษานี้จะทำการวัดด้วยคะแนน Accuracy, Precision, Recall และ F1-score ซึ่งสามารถคำนวณได้ดังนี้

3.2.5.1 คะแนน Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

โดยที่ **TP : True Positive** คือ จำนวนวันที่ทำนายผลว่าเป็นแนวโน้มราคาขึ้น (Up) แล้วราคามีแนวโน้มขึ้นตามจริง (Up)

TN : True Negative คือ จำนวนวันที่ทำนายผลว่าเป็นแนวโน้มราคาลง (Down) แล้วราคามีแนวโน้มลงตามจริง (Down)

FP : False Positive คือ จำนวนวันที่ทำนายผลว่าเป็นแนวโน้มราคาขึ้น (Up) แต่ราคาจริงมีแนวโน้มลง (Down)

FN : False Negative คือ จำนวนวันที่ทำนายผลว่าเป็นแนวโน้มราคาลง (Down) แต่ราคาจริงมีแนวโน้มขึ้น (Up)

3.2.5.2 คะแนน Precision

$$Precision = \frac{TP}{TP + FP}$$

3.2.5.3 คะแนน Recall

$$Recall = \frac{TP}{TP + FN}$$

3.2.5.4 คะแนน F1-score

$$F1 = \frac{2 * Precision * recall}{Precision + recall}$$

3.2.6 ขั้นตอนการปรับค่าพารามิเตอร์ (Parameter tuning)

เป็นขั้นตอนสำหรับการกำหนดค่าพารามิเตอร์ที่เกี่ยวข้องในขั้นตอนการทำนายผลเพื่อหาค่าหรือจำนวนที่เหมาะสมที่ทำให้วิธีการทำนายผลมีประสิทธิภาพดีที่สุด โดยในการศึกษานี้จะมีค่าพารามิเตอร์ที่ทำการพิจารณาทั้งหมด 2 ตัว ได้แก่ ขนาดของรูปแบบ (Window size) และจำนวนรูปแบบ (Number of patterns)

3.2.6.1 ขนาดของรูปแบบ (Window size)

ขนาดของรูปแบบ คือ ค่าขนาดของลำดับย่อยอนุกรมเวลา หรือจำนวนวันของรูปแบบที่กำหนด เป็นค่าที่สำคัญในการกำหนดกรอบวันของราคาที่จะทำการศึกษานบนหลักทรัพย์แบบรายวัน

3.2.6.2 จำนวนรูปแบบ (Number of patterns)

จำนวนรูปแบบ คือ ค่าของจำนวนรูปแบบทั้งหมดที่ได้รับมาจากในขั้นตอนเทคนิคการทำเหมืองข้อมูลผ่านวิธีค้นพบโมทีฟในอนุกรมเวลา ซึ่งจะนำรูปแบบทั้งหมดที่ได้มาใช้เป็นตัวแปรในการทำนายผลผ่านวิธีการที่ได้ทดลองศึกษา

3.2.7 ขั้นตอนการวิเคราะห์แนวโน้มหลักทรัพย์รายวันผ่านแพลตฟอร์มซื้อขายจำลอง

เป็นขั้นตอนในการทดสอบสมมุติฐานจากผลการทดลองจากผลการวิเคราะห์ทางสถิติและผลการทำนายจากโมเดลวิธีการเรียนรู้ของเครื่อง โดยแพลตฟอร์มที่เลือกใช้คือแอปพลิเคชัน Streaming Click2Win



ภาพที่ 3.8 แพลตฟอร์ม Streaming Click2Win

บทที่ 4

ผลการดำเนินงาน

4.1 การเก็บข้อมูล

ทำการเก็บรวบรวมชุดข้อมูลตลาดหลักทรัพย์รายวันจากตลาดหลักทรัพย์แห่งประเทศไทย (SET) ย้อนหลังเป็นเวลา 5 ปี นับตั้งแต่ 10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565 ในกลุ่มอุตสาหกรรมที่แตกต่างกัน

หุ้น (ชื่อย่อ)	เวลาที่จัดเก็บ (เริ่มต้น ถึง ล่าสุด)	ขนาดข้อมูล
KBANK.BK	10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565	1,214 แถว
PTTEP.BK	10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565	1,214 แถว
CPALL.BK	10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565	1,214 แถว
DELTA.BK	10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565	1,214 แถว
SCC.BK	10 พฤศจิกายน 2560 ถึง 10 พฤศจิกายน 2565	1,214 แถว

ตารางที่ 4.1 รายการข้อมูลตลาดหลักทรัพย์ที่ใช้ในศึกษา

4.2 ผลการทดลองของเทคนิควิธีการทำเหมืองข้อมูล (Data mining techniques)

4.2.1 การค้นพบโมทีฟในอนุกรมเวลา (Time series motif discovery)

4.2.1.1 การวัดระยะห่างความคล้ายคลึงกัน (Similarity distance measure)

ใช้อัลกอริทึม Dynamic Time Warping ในการค้นหารูปแบบในอดีตที่มีลักษณะคล้ายคลึงกับรูปแบบปัจจุบันบนหลักทรัพย์ในแต่ละรายการที่ได้มีการจัดเก็บลำดับย่อยของชุดข้อมูลและแนวโน้มของราคาในวันถัดไป โดยในแต่ละรอบจะมีการกำหนดค่าพารามิเตอร์ที่แตกต่างกัน ดังนี้

Window size = [5, 6, 7, 8, 9, 10]

Number of patterns = [5, 10, 15, 20, 25, 30]

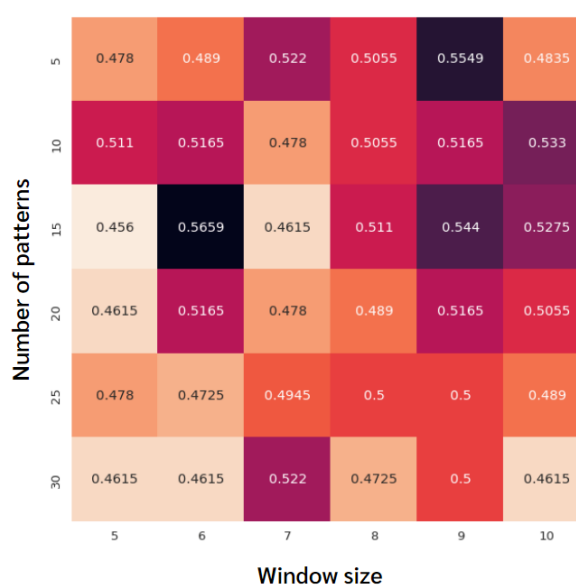
4.2.2 หลักทรัพย์ KBANK.BK (กลุ่มธุรกิจการเงิน)

4.2.2.1 วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

(1) The best parameter on KBANK.BK

จำนวนรูปแบบ (Number of patterns : N) = 15 patterns

ขนาดรูปแบบ (Window size : W) = 6 days



ภาพที่ 4.1 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ KBANK.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 15 และ W = 6 จะได้

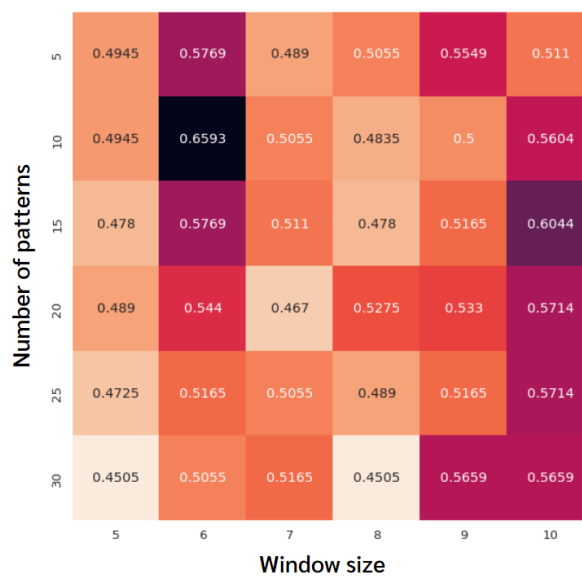
Accuracy	เท่ากับ 0.5659 หรือ 56.59%
Precision	เท่ากับ 0.6000 หรือ 60.00%
Recall	เท่ากับ 0.6061 หรือ 60.61%
F1-score	เท่ากับ 0.6030 หรือ 60.30%

4.2.2.2 วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

(1) The best parameter on KBANK.BK

จำนวนรูปแบบ (Number of patterns : N) = 10 patterns

ขนาดรูปแบบ (Window size : W) = 6 days



ภาพที่ 4.2 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ KBANK.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 10 และ W = 6 จะได้

Accuracy เท่ากับ 0.6593 หรือ 65.93%

Precision เท่ากับ 0.7079 หรือ 70.79%

Recall เท่ากับ 0.6364 หรือ 63.64%

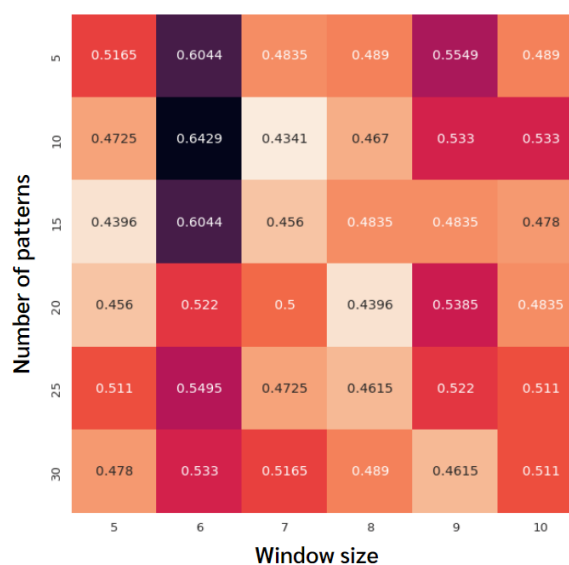
F1-score เท่ากับ 0.6702 หรือ 67.02%

4.2.2.3 วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best parameter on KBANK.BK

จำนวนรูปแบบ (Number of patterns : N) = 10 patterns

ขนาดรูปแบบ (Window size : W) = 6 days



ภาพที่ 4.3 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ KBANK.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 10 และ W = 6 จะได้

Accuracy เท่ากับ 0.6429 หรือ 64.29%

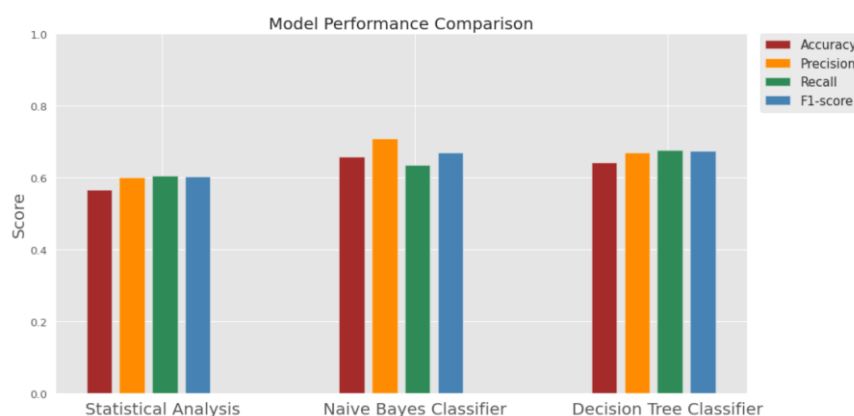
Precision เท่ากับ 0.6700 หรือ 67.00%

Recall เท่ากับ 0.6768 หรือ 67.68%

F1-score เท่ากับ 0.6734 หรือ 67.34%

4.2.2.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

ประสิทธิภาพของแต่ละวิธีในการทำนายแนวโน้มราคาหลักทรัพย์ KBANK.BK



ภาพที่ 4.4 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ KBANK.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Statistical Analysis	0.5659	0.6000	0.6061	0.6030
Naïve Bayes Classifier	0.6593	0.7079	0.6364	0.6702
Decision Tree Classifier	0.6429	0.6700	0.6768	0.6734

ตารางที่ 4.2 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ KBANK.BK

จากผลการทดลองจากทั้งสามวิธีในการทำนายผลบนชุดข้อมูลหลักทรัพย์ KBANK.BK จากกลุ่มธุรกิจการเงิน จะเห็นได้ว่ากรอบรูปแบบราคาในระยะเวลา 6 วัน (ขนาด Window size เท่ากับ 6) ให้ค่า Accuracy ได้ดีที่สุดจากทั้งสามวิธีการทำนายแนวโน้มผ่านเทคนิคการทำเหมืองข้อมูล โดยวิธีจำแนกแบบเบย์อย่างง่ายเป็นวิธีที่ให้ค่า Accuracy สูงที่สุด คือ 0.6593 และวิธีจำแนกต้นไม้การตัดสินใจเป็นวิธีที่ให้ค่า F1-score สูงที่สุด คือ 0.6734

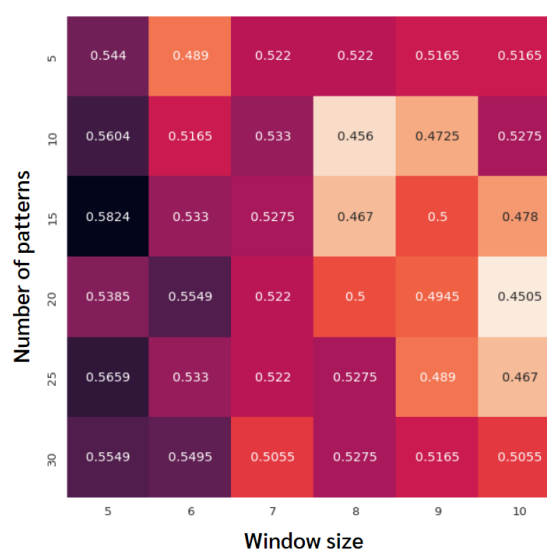
4.2.3 หลักทรัพย์ PTTEP.BK (กลุ่มทรัพยากร)

4.2.3.1 วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

(1) The best parameter on PTTEP.BK

จำนวนรูปแบบ (Number of pattern : N) = 15 patterns

ขนาดรูปแบบ (Window size : W) = 5 days



ภาพที่ 4.5 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ PTTEP.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 15 และ W = 5 จะได้

Accuracy เท่ากับ 0.5824 หรือ 58.24%

Precision เท่ากับ 0.6220 หรือ 62.20%

Recall เท่ากับ 0.7383 หรือ 73.83%

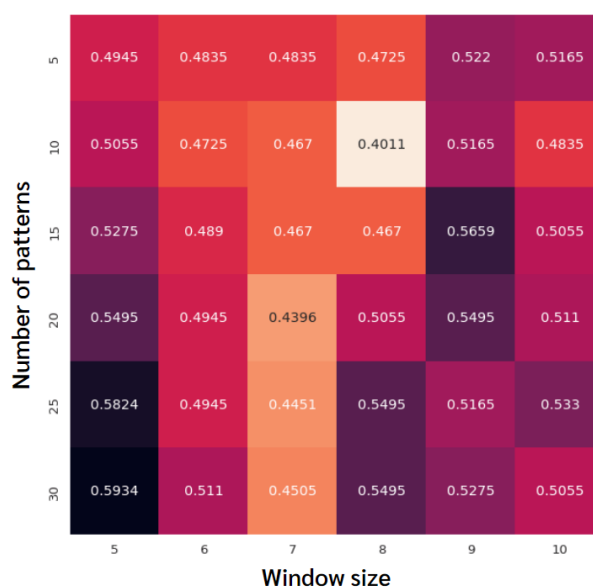
F1-score เท่ากับ 0.6752 หรือ 67.52%

4.2.3.2 วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

(1) The best parameter on PTTEP.BK

จำนวนรูปแบบ (Number of pattern : N) = 30 patterns

ขนาดรูปแบบ (Window size : W) = 5 days



ภาพที่ 4.6 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ PTTEP.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 30 และ W = 5 จะได้

Accuracy เท่ากับ 0.5934 หรือ 59.34%

Precision เท่ากับ 0.6387 หรือ 63.87%

Recall เท่ากับ 0.7103 หรือ 71.03%

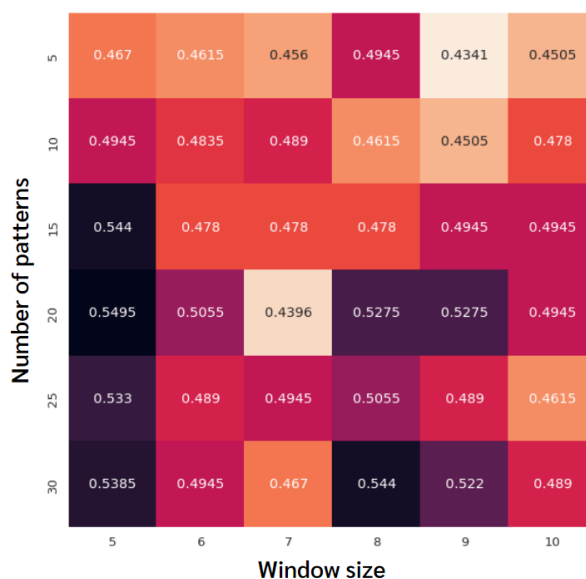
F1-score เท่ากับ 0.6726 หรือ 67.26%

4.2.3.3 วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best parameter on PTTEP.BK

จำนวนรูปแบบ (Number of pattern : N) = 20 patterns

ขนาดรูปแบบ (Window size : W) = 5 days



ภาพที่ 4.7 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ PTTEP.BK

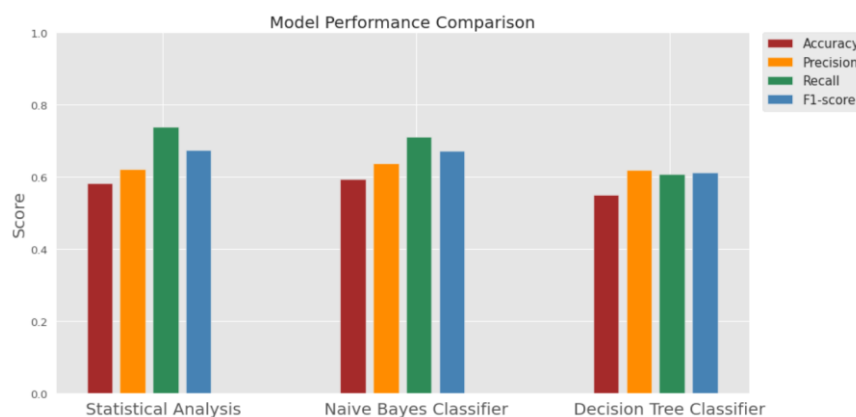
(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 20 และ W = 5 จะได้

Accuracy	เท่ากับ 0.5495 หรือ 54.95%
Precision	เท่ากับ 0.6122 หรือ 61.22%
Recall	เท่ากับ 0.5607 หรือ 56.07%
F1-score	เท่ากับ 0.5854 หรือ 58.54%

4.2.3.1 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

ประสิทธิภาพของแต่ละวิธีในการทำนายแนวโน้มราคาหลักทรัพย์ PTTEP.BK



ภาพที่ 4.8 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ PTTEP.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Statistical Analysis	0.5824	0.6220	0.7383	0.6752
Naïve Bayes Classifier	0.5934	0.6387	0.7103	0.6726
Decision Tree Classifier	0.5495	0.6190	0.6075	0.6132

ตารางที่ 4.3 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ PTTEP.BK

จากผลการทดลองจากทั้งสามวิธีในการทำนายผลบนชุดข้อมูลหลักทรัพย์ PTTEP.BK จากกลุ่มทรัพยากร จะเห็นได้ว่ากรอบรูปแบบราคาในระยะเวลา 5 วัน (ขนาด Window size เท่ากับ 5) ให้ค่า Accuracy ได้ดีที่สุดจากทั้งสามวิธีการทำนายแนวโน้มผ่านเทคนิคการทำเหมืองข้อมูล โดยวิธีจำแนกแบบเบย์อย่างง่ายเป็นวิธีที่ให้ค่า Accuracy สูงที่สุด คือ 0.5934 และวิธีการวิเคราะห์ทางสถิติเป็นวิธีที่ให้ค่า F1-score สูงที่สุด คือ 0.6752

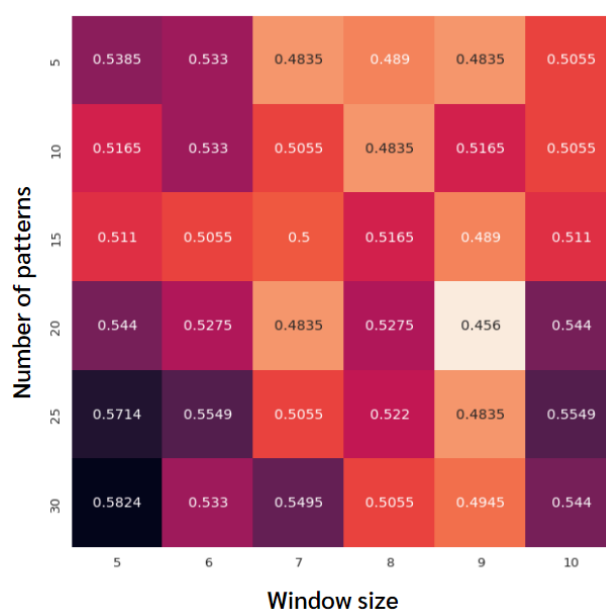
4.2.4 หลักทรัพย์ CPALL.BK (กลุ่มบริการ)

4.2.4.1 วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

(1) The best parameter on CPALL.BK

จำนวนรูปแบบ (Number of pattern : N) = 30 patterns

ขนาดรูปแบบ (Window size : W) = 5 days



ภาพที่ 4.9 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด

ค่าพารามิเตอร์สำหรับการวิเคราะห์ทางสถิติบนหลักทรัพย์ CPALL.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 30 และ W = 5 จะได้

Accuracy เท่ากับ 0.5824 หรือ 58.24%

Precision เท่ากับ 0.6084 หรือ 60.84%

Recall เท่ากับ 0.8131 หรือ 81.31%

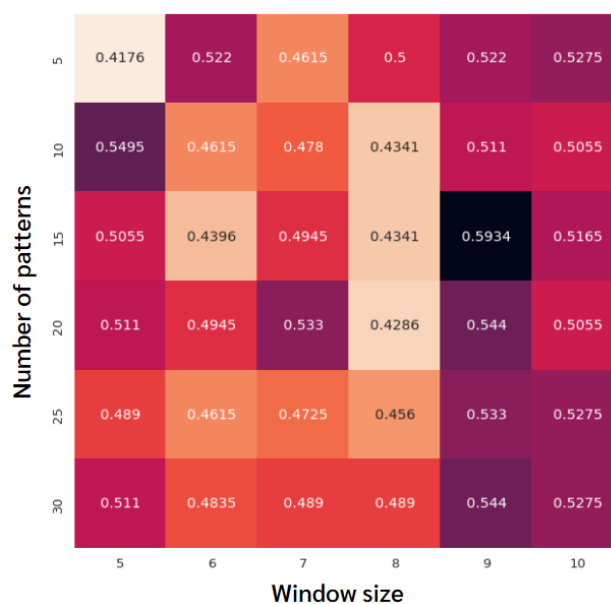
F1-score เท่ากับ 0.6960 หรือ 69.60%

4.2.4.2 วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

(1) The best parameter on CPALL.BK

จำนวนรูปแบบ (Number of pattern : N) = 15 patterns

ขนาดรูปแบบ (Window size : W) = 9 days



ภาพที่ 4.10 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ CPALL.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 15 และ W = 9 จะได้

Accuracy เท่ากับ 0.5934 หรือ 59.34%

Precision เท่ากับ 0.6634 หรือ 66.34%

Recall เท่ากับ 0.6262 หรือ 62.62%

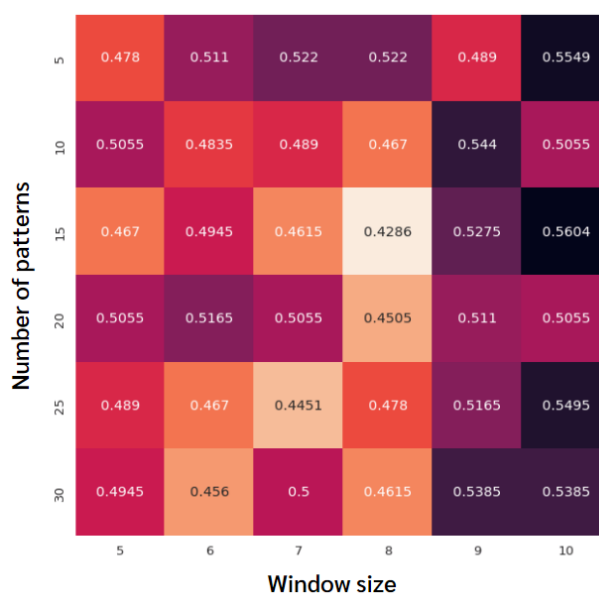
F1-score เท่ากับ 0.6442 หรือ 64.42%

4.2.4.3 วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best parameter on CPALL.BK

จำนวนรูปแบบ (Number of pattern : N) = 15 patterns

ขนาดรูปแบบ (Window size : W) = 10 days



ภาพที่ 4.11 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ CPALL.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 15 และ W = 10 จะได้

Accuracy เท่ากับ 0.5604 หรือ 56.04%

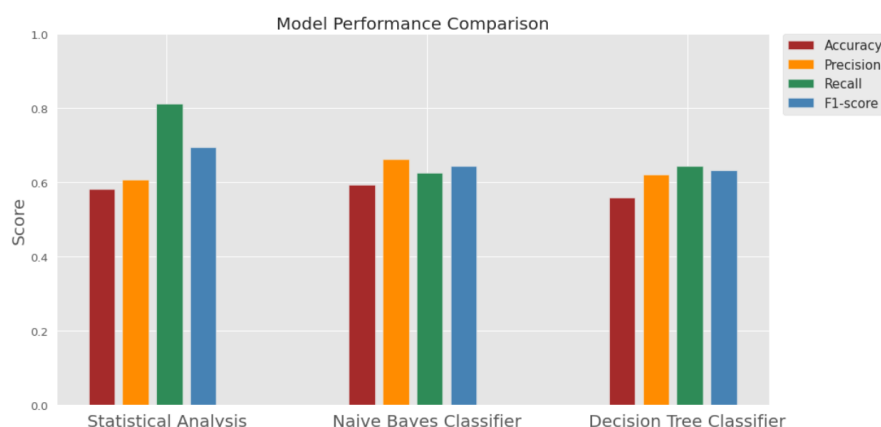
Precision เท่ากับ 0.6216 หรือ 62.16%

Recall เท่ากับ 0.6449 หรือ 64.49%

F1-score เท่ากับ 0.6330 หรือ 63.30%

4.2.4.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

ประสิทธิภาพของแต่ละวิธีในการทำนายแนวโน้มราคาหลักทรัพย์ CPALL.BK



ภาพที่ 4.12 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ CPALL.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Statistical Analysis	0.5824	0.6084	0.8131	0.6960
Naïve Bayes Classifier	0.5934	0.6634	0.6262	0.6442
Decision Tree Classifier	0.5604	0.6216	0.6449	0.6330

ตารางที่ 4.4 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ CPALL.BK

จากผลการทดลองจากทั้งสามวิธีในการทำนายผลบนชุดข้อมูลหลักทรัพย์ CPALL.BK จะเห็นได้ว่ากรอบรูปแบบราคาที่ให้ค่า Accuracy สูงที่สุดนั้นค่อนข้างหลากหลายขึ้นอยู่กับแต่ละวิธีที่เลือกใช้ในการทำนาย ซึ่งเมื่อพิจารณาภาพรวมจะพบว่าวิธีจำแนกแบบเบย์อย่างง่ายหลังปรับค่าพารามิเตอร์เป็นวิธีที่ให้ค่า Accuracy สูงที่สุด คือ 0.5934 และวิธีการวิเคราะห์ทางสถิติหลังปรับค่าพารามิเตอร์เป็นวิธีที่ให้ค่า F1-score สูงที่สุด คือ 0.6960

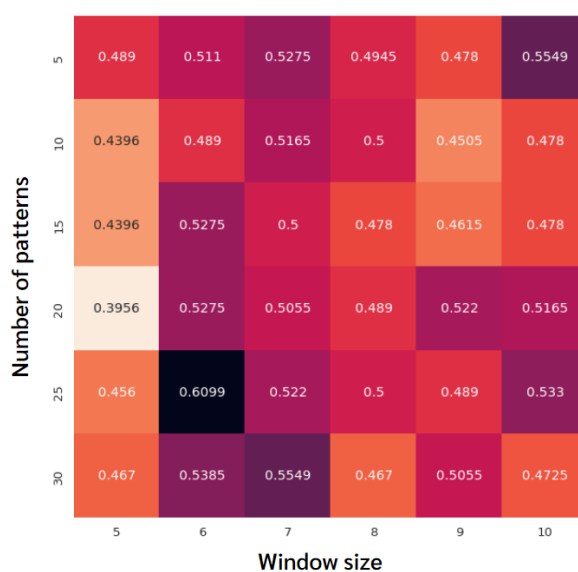
4.2.5 หลักทรัพย์ DELTA.BK (กลุ่มเทคโนโลยี)

4.2.5.1 วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

(1) The best parameter on DELTA.BK

จำนวนรูปแบบ (Number of pattern : N) = 25 patterns

ขนาดรูปแบบ (Window size : W) = 6 days



ภาพที่ 4.13 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ DELTA.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 25 และ W = 6 จะได้

Accuracy เท่ากับ 0.6099 หรือ 60.99%

Precision เท่ากับ 0.5818 หรือ 58.18%

Recall เท่ากับ 0.7191 หรือ 71.91%

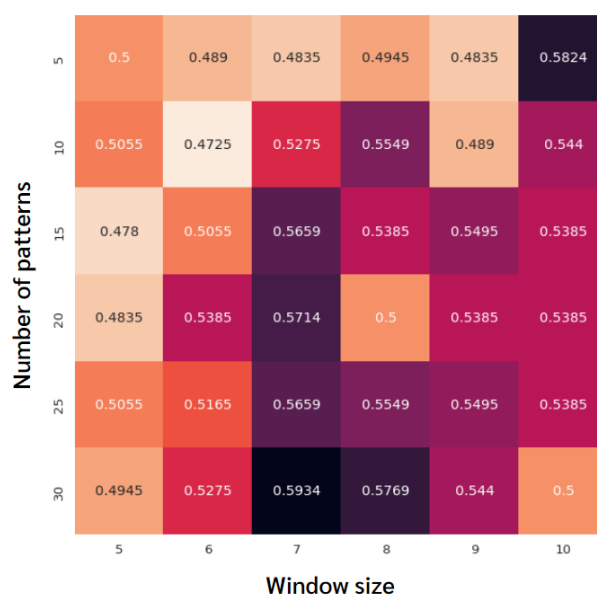
F1-score เท่ากับ 0.6432 หรือ 64.32%

4.2.5.2 วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

(1) The best parameter on DELTA.BK

จำนวนรูปแบบ (Number of pattern : N) = 30 patterns

ขนาดรูปแบบ (Window size : W) = 7 days



ภาพที่ 4.14 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนดค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทรัพย์ DELTA.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 30 และ W = 7 จะได้

Accuracy เท่ากับ 0.5934 หรือ 59.34%

Precision เท่ากับ 0.5773 หรือ 57.73%

Recall เท่ากับ 0.6292 หรือ 62.92%

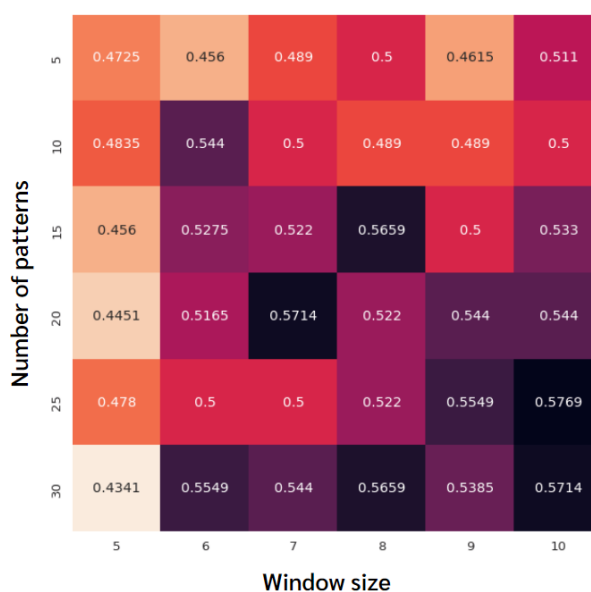
F1-score เท่ากับ 0.6022 หรือ 60.22%

4.2.5.3 วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best parameter on DELTA.BK

จำนวนรูปแบบ (Number of pattern : N) = 25 patterns

ขนาดรูปแบบ (Window size : W) = 10 days



ภาพที่ 4.15 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ DELTA.BK

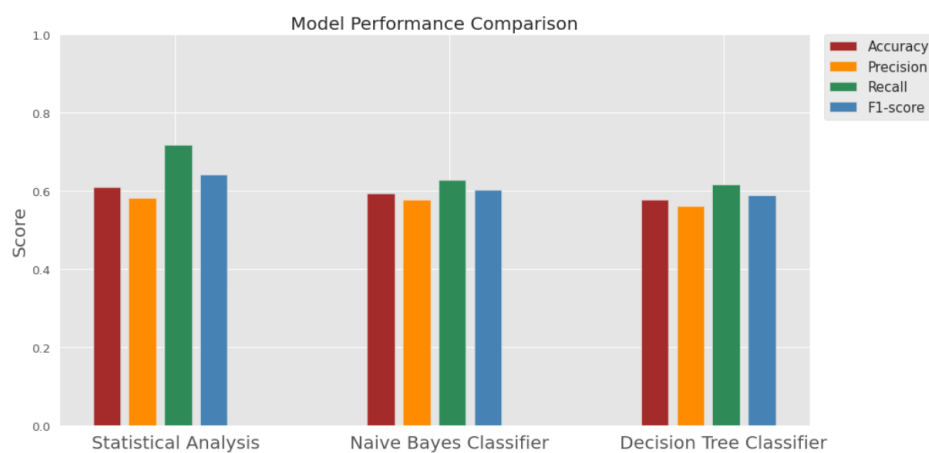
(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 25 และ W = 10 จะได้

Accuracy	เท่ากับ 0.5769 หรือ 57.69%
Precision	เท่ากับ 0.5612 หรือ 56.12%
Recall	เท่ากับ 0.6180 หรือ 61.80%
F1-score	เท่ากับ 0.5882 หรือ 58.82%

4.2.5.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

ประสิทธิภาพของแต่ละวิธีในการทำนายแนวโน้มราคาหลักทรัพย์ DELTA.BK



ภาพที่ 4.16 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ DELTA.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Statistical Analysis	0.6099	0.5818	0.7191	0.6432
Naïve Bayes Classifier	0.5934	0.5773	0.6292	0.6022
Decision Tree Classifier	0.5769	0.5612	0.6180	0.5882

ตารางที่ 4.5 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ DELTA.BK

จากผลการทดลองจากทั้งสามวิธีในการทำนายผลบนชุดข้อมูลหลักทรัพย์ DELTA.BK จะเห็นได้ว่ากรอบรูปแบบราคาที่ให้ค่า Accuracy สูงที่สุดนั้นค่อนข้างหลากหลายขึ้นอยู่กับแต่ละวิธีที่เลือกใช้ในการทำนาย ซึ่งเมื่อพิจารณาภาพรวมจะพบว่าวิธีการวิเคราะห์ทางสถิติหลังปรับค่าพารามิเตอร์เป็นวิธีที่ให้ค่า Accuracy และ F1-score สูงที่สุด คือ 0.6099 และ 0.6432 ตามลำดับ

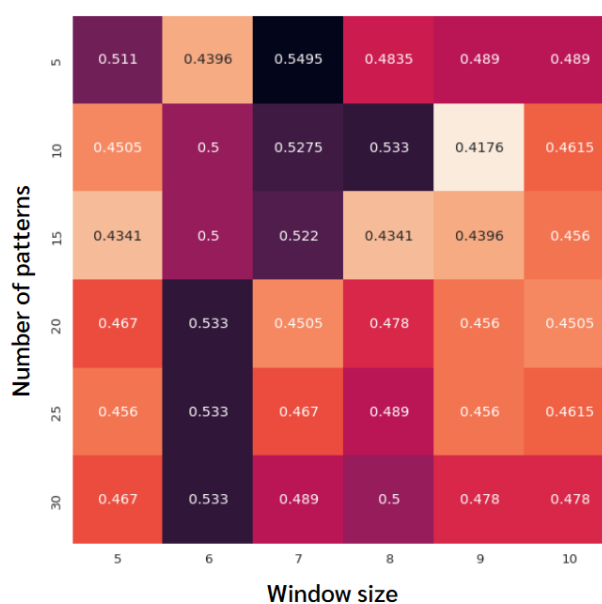
4.2.6 หลักทรัพย์ SCC.BK (กลุ่มอสังหาริมทรัพย์และการก่อสร้าง)

4.2.6.1 วิธีการวิเคราะห์ทางสถิติ (Statistical Analysis)

(1) The best parameter on SCC.BK

จำนวนรูปแบบ (Number of pattern : N) = 5 patterns

ขนาดรูปแบบ (Window size : W) = 7 days



ภาพที่ 4.17 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีการวิเคราะห์ทางสถิติบนหลักทรัพย์ SCC.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 5 และ W = 7 จะได้

Accuracy เท่ากับ 0.5495 หรือ 54.95%

Precision เท่ากับ 0.5567 หรือ 55.67%

Recall เท่ากับ 0.5806 หรือ 58.06%

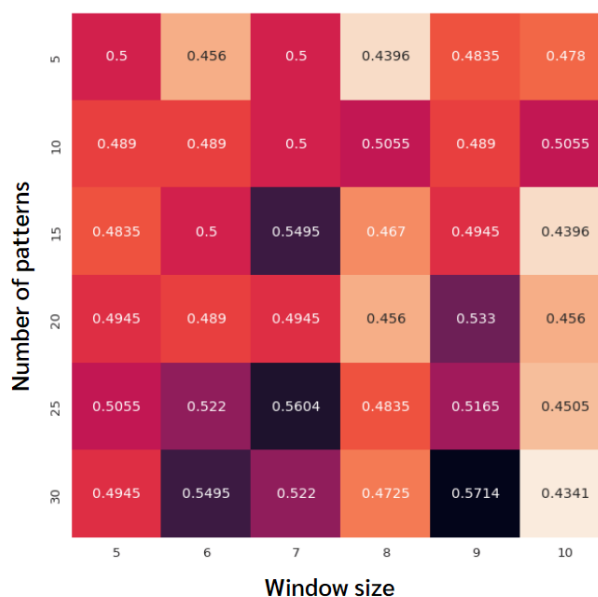
F1-score เท่ากับ 0.5684 หรือ 56.84%

4.2.6.2 วิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

(1) The best parameter on SCC.BK

จำนวนรูปแบบ (Number of pattern : N) = 30 patterns

ขนาดรูปแบบ (Window size : W) = 9 days



ภาพที่ 4.18 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกแบบเบย์อย่างง่ายบนหลักทฤษฎี SCC.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 30 และ W = 9 จะได้

Accuracy เท่ากับ 0.5714 หรือ 57.14%

Precision เท่ากับ 0.5862 หรือ 58.62%

Recall เท่ากับ 0.5439 หรือ 54.39%

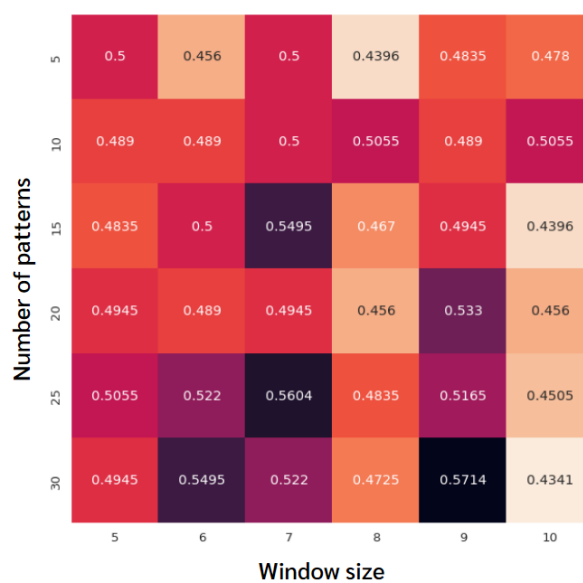
F1-score เท่ากับ 0.5667 หรือ 56.67%

4.2.6.3 วิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best parameter on SCC.BK

จำนวนรูปแบบ (Number of pattern : N) = 30 patterns

ขนาดรูปแบบ (Window size : W) = 9 days



ภาพที่ 4.19 แผนภาพ Heatmap แสดงค่า Accuracy ในแต่ละรอบการกำหนด
ค่าพารามิเตอร์สำหรับวิธีจำแนกต้นไม้การตัดสินใจบนหลักทรัพย์ SCC.BK

(2) Evaluation Metrics

เมื่อรันผลโดยกำหนด N = 30 และ W = 9 จะได้

Accuracy	เท่ากับ 0.5714 หรือ 57.14%
Precision	เท่ากับ 0.5862 หรือ 58.62%
Recall	เท่ากับ 0.5439 หรือ 54.39%
F1-score	เท่ากับ 0.5667 หรือ 56.67%

4.2.6.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

ประสิทธิภาพของแต่ละวิธีในการทำนายแนวโน้มราคาหลักทรัพย์ SCC.BK



ภาพที่ 4.20 กราฟแสดงประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ SCC.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Statistical Analysis	0.5495	0.5567	0.5806	0.5684
Naïve Bayes Classifier	0.5714	0.5862	0.5484	0.5667
Decision Tree Classifier	0.5604	0.5684	0.5806	0.5745

ตารางที่ 4.6 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ SCC.BK

จากผลการทดลองจากทั้งสามวิธีในการทำนายผลบนชุดข้อมูลหลักทรัพย์ SCC.BK จะเห็นได้ว่าการออกแบบราคาที่ให้ค่า Accuracy สูงที่สุดนั้นค่อนข้างหลากหลายขึ้นอยู่กับแต่ละวิธีที่เลือกใช้ในการทำนาย ซึ่งเมื่อพิจารณาภาพรวมจะพบว่าวิธีจำแนกแบบเบย์อย่างง่ายหลังปรับค่าพารามิเตอร์เป็นวิธีที่ให้ค่า Accuracy สูงที่สุด คือ 0.5714 และวิธีจำแนกต้นไม้การตัดสินใจหลังปรับค่าพารามิเตอร์เป็นวิธีที่ให้ค่า F1-score สูงที่สุด คือ 0.5745

4.3 ผลการทดลองของวิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค

4.3.1 การคำนวณตัวชี้วัดทางเทคนิค (Technical Indicators)

สำหรับวิธีการดึงคุณลักษณะโดยใช้ตัวชี้วัดทางเทคนิคเป็นตัวแปรนำเข้าในการสร้างโมเดลทำนายผล จะใช้ไลบรารี Talibs ในการดึงคุณลักษณะที่เป็นอินดิเคเตอร์ผ่านการคำนวณจากตัวแปรราคาและปริมาณซื้อขายรายวันที่ได้รับมาจากชุดข้อมูลดิบ ซึ่งตัวไลบรารีสามารถคำนวณอินดิเคเตอร์ได้ทั้งสิ้นจำนวน 84 ตัว โดยสามารถจัดกลุ่มอินดิเคเตอร์ออกเป็นทั้งหมด 4 กลุ่มใหญ่ ได้แก่

1) Trend Indicator	เป็นกลุ่มอินดิเคเตอร์สำหรับบอกแนวโน้มซึ่งจะช่วยระบุทิศทางหรือแนวโน้มของราคา โดยเปรียบเทียบข้อมูลราคาในช่วงเวลาต่างๆ
2) Volume Indicator	เป็นกลุ่มอินดิเคเตอร์ที่ใช้วิเคราะห์ปริมาณการซื้อขาย ซึ่งจะช่วยยืนยันแนวโน้มและรูปแบบของราคา
3) Momentum Indicator	เป็นกลุ่มอินดิเคเตอร์สำหรับใช้วัดอัตราการเปลี่ยนแปลงหรือความเร็วของการเคลื่อนไหวของราคา
4) Volatility Indicator	เป็นกลุ่มอินดิเคเตอร์ที่ใช้วัดความผันผวนของตลาด โดยแยกช่วงราคาทั้งหมดในช่วงเวลาที่ระบุ

ตารางที่ 4.7 ประเภทของอินดิเคเตอร์

โดยอินดิเคเตอร์ที่คำนวณได้ทั้งหมดในแต่ละกลุ่มจะนำมาใช้เป็นตัวแปรนำเข้าสำหรับการเทรนโมเดลเพื่อทำนายผล โดยมีคลาสเป้าหมาย คือ แนวโน้มของราคาในวันถัดไปซึ่งประกอบด้วย แนวโน้มขาขึ้น (+1) และแนวโน้มขาลง (-1) โดยสามารถกำหนดข้อมูลในคลาสเป้าหมายได้จากการพิจารณาราคาปิดของหลักทรัพย์ในเวลาปัจจุบัน เทียบกับราคาปิดของหลักทรัพย์ในวันก่อนหน้า

4.3.2 การกำหนดค่าไฮเปอร์พารามิเตอร์และเทคนิคการดึงคุณลักษณะของโมเดล (Hyperparameter tuning and Feature extraction techniques)

กำหนดเทคนิคการดึงคุณลักษณะในการเทรนโมเดลวิธีจำแนกแบบเบย์อย่างง่าย

Method	Feature extraction techniques
GNB	Gaussian Naïve Bayes
GNB_MinMax	Gaussian Naïve Bayes and MinMax scaling
GNB_MinMax_PCA	Gaussian Naïve Bayes and MinMax scaling and Principal Component Analysis (PCA)
GNB_MinMax_FA	Gaussian Naïve Bayes and MinMax scaling and Factor Analysis (FA)
GNB_LDA	Gaussian Naïve Bayes and Linear Discriminant Analysis (LDA)

ตารางที่ 4.8 เทคนิควิธีการดึงคุณลักษณะต่างๆบนวิธีจำแนกแบบเบย์อย่างง่าย

กำหนดค่าไฮเปอร์พารามิเตอร์ที่ทำการทดลองปรับในการเทรนโมเดลวิธีจำแนกต้นไม้การตัดสินใจ

criterion	['gini', 'entropy']
max_depth	[3, 5, 7, 9, 10, 12, 15, 20, 25, 30]
min_samples_leaf	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

ตารางที่ 4.9 ค่าไฮเปอร์พารามิเตอร์บนโมเดลวิธีจำแนกต้นไม้การตัดสินใจ

กำหนดค่าไฮเปอร์พารามิเตอร์ที่ทำการทดลองปรับในการเทรนโมเดลโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น

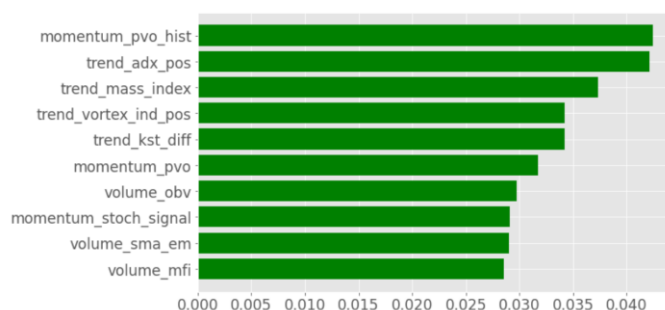
activation	['identity', 'logistic', 'tanh', 'relu']
hidden_layer_sizes	[(10,), (20,), (30,), (40,), (50,)]
alpha	[0.0001, 0.001, 0.01, 0.05]

ตารางที่ 4.10 ค่าไฮเปอร์พารามิเตอร์บนโมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น

4.3.1 การคัดเลือกฟีเจอร์ (Feature selection)

ทำการคัดเลือกตัวแปรเพื่อเลือกใช้อินดิเคเตอร์สำคัญที่มีผลต่อคลาสเป้าหมาย นั่นก็คือแนวโน้มของราคาหลักทรัพย์ โดยสามารถหาได้จากการคำนวณหาค่า Features Importance ของตัวแปรทั้งหมดผ่านฟังก์ชันของการเทรนโมเดลต้นไม้การตัดสินใจ (Decision Tree)

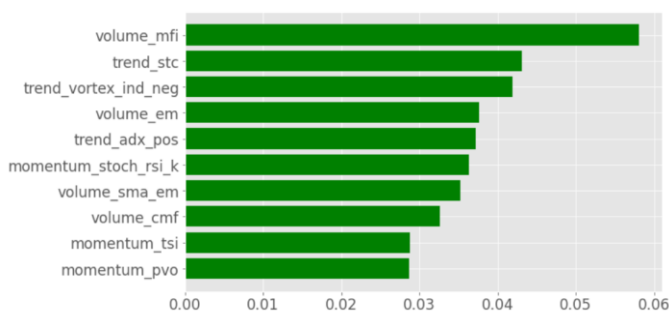
4.3.1.1 Feature selection on KBANK.BK stock dataset



ภาพที่ 4.21 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ KBANK.BK

จากภาพจะเห็นได้ว่าอินดิเคเตอร์กลุ่ม Momentum, Trend และ Volume Indicators มีความสำคัญต่อแนวโน้มของราคาหลักทรัพย์ KBANK.BK ตามลำดับ อย่างไรก็ตามเมื่อพิจารณาจากภาพรวมจะเห็นได้ว่าอินดิเคเตอร์แต่ละตัวยังคงมีความสัมพันธ์กับคลาสเป้าหมายค่อนข้างต่ำ โดยตัวแปรอินดิเคเตอร์ Percentage Volume Oscillator (PVO) จากกลุ่ม Momentum Indicator มีค่า Importance ที่สูงสุดอยู่ที่ 0.0425

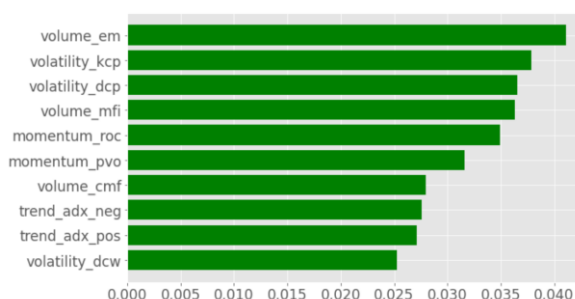
4.3.1.2 Feature selection on PTTEP.BK stock dataset



ภาพที่ 4.22 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ PTTEP.BK

จากภาพจะเห็นได้ว่าอินดิเคเตอร์กลุ่ม Volume, Trend และ Momentum มีความสำคัญต่อแนวโน้มของราคาหลักทรัพย์ PTTEP.BK ตามลำดับแต่เมื่อพิจารณาภาพรวมจะเห็นได้ว่าอินดิเคเตอร์แต่ละตัวยังคงมีความสัมพันธ์กับคลาสเป้าหมายค่อนข้างต่ำเช่นเดียวกัน โดยอินดิเคเตอร์ Money Flow Index (MFI) จากกลุ่ม Volume Indicator มีค่า Importance สูงที่สุดกว่าอินดิเคเตอร์อื่นๆ โดยมีค่าอยู่ที่ 0.0581

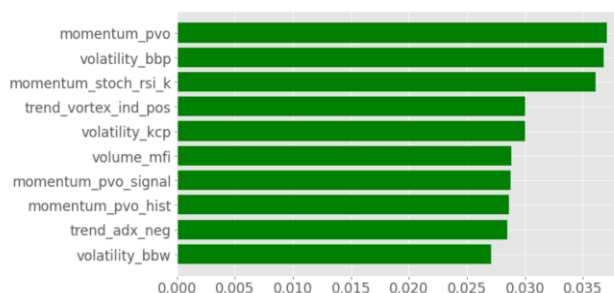
4.3.1.3 Feature selection on CPALL.BK stock dataset



ภาพที่ 4.23 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ CPALL.BK

จากภาพจะเห็นได้ว่าอินดิเคเตอร์ในทุกๆกลุ่ม มีความสำคัญต่อแนวโน้มของราคาหลักทรัพย์ CPALL.BK ไล่เรียงกัน แต่เมื่อพิจารณาภาพรวมจะเห็นได้ว่าอินดิเคเตอร์แต่ละตัวยังคงมีความสัมพันธ์กับคลาสเป้าหมายค่อนข้างต่ำเช่นเดียวกัน โดยตัวแปรอินดิเคเตอร์ EM จากกลุ่ม Volume Indicator มีค่า Importance สูงที่สุดอยู่ที่ 0.0411

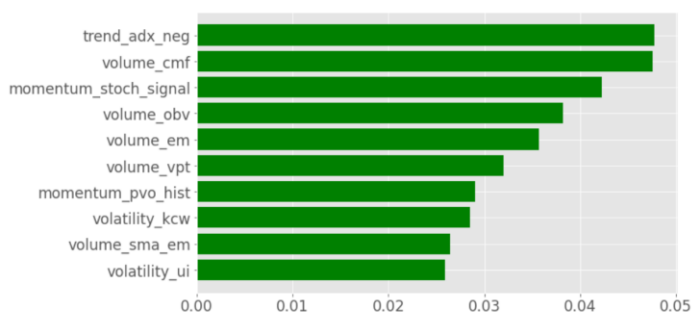
4.3.1.4 Features selection on DELTA.BK stock dataset



ภาพที่ 4.24 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ DELTA.BK

จากภาพจะเห็นได้ว่าอินดิเคเตอร์ในทุกๆกลุ่ม มีความสำคัญต่อแนวโน้มของราคาหลักทรัพย์ DELTA.BK ไล่เรียงกัน แต่เมื่อพิจารณาภาพรวมจะเห็นได้ว่าอินดิเคเตอร์แต่ละตัวยังคงมีความสัมพันธ์กับคลาสเป้าหมายค่อนข้างต่ำเช่นเดียวกัน โดยอินดิเคเตอร์ Percentage Volume Oscillator (PVO) จากกลุ่ม Momentum Indicator มีค่า Importance สูงสุดอยู่ที่ 0.0371

4.3.1.5 Feature selection on SCC.BK stock dataset

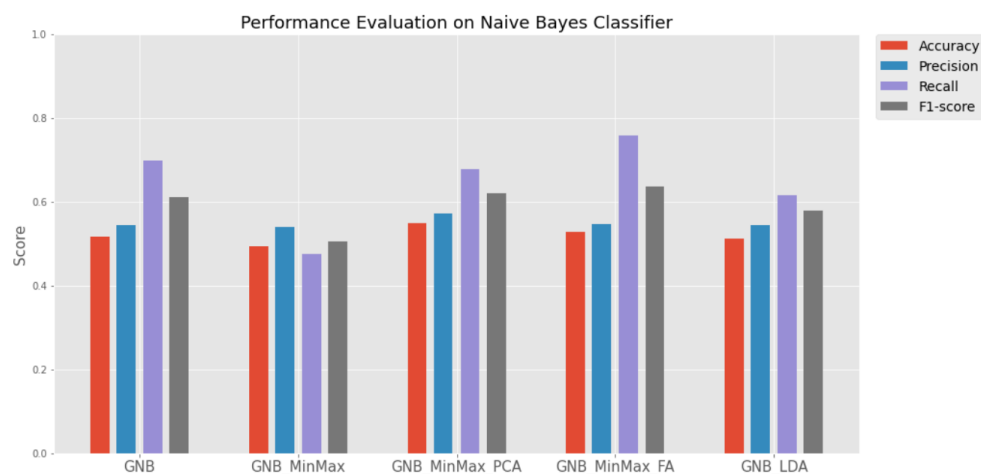


ภาพที่ 4.25 กราฟแสดงลำดับตัวแปรที่มีค่า Importance สูงที่สุดบนหลักทรัพย์ SCC.BK

จากภาพจะเห็นได้ว่าอินดิเคเตอร์กลุ่มในทุกๆกลุ่ม มีความสำคัญต่อแนวโน้มของราคาหลักทรัพย์ SCC.BK ไล่เรียงกัน อย่างไรก็ตามเมื่อพิจารณาภาพรวมจะเห็นได้ว่าอินดิเคเตอร์แต่ละตัวยังคงมีความสัมพันธ์กับคลาสเป้าหมายค่อนข้างต่ำเช่นเดียวกัน โดยตัวแปรอินดิเคเตอร์ Negative Directional Indicator (-DI) จากกลุ่ม Trend Indicator มีค่า Importance สูงที่สุดอยู่ที่ 0.0478

4.3.2 หลักทรัพย์ KBANK.BK (กลุ่มธุรกิจการเงิน)

4.3.2.1 โมเดลวิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 4.26 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างควบคู่กับเทคนิควิธีที่ต่างกันบนหลักทรัพย์ KBANK.BK

(1) The best combination techniques on GNB Model :

Gaussian Naïve Bayes and MinMax-scaling and PCA

(2) Evaluation Metrics

เมื่อรันผลผ่านโมเดล GNB_MinMax_PCA จะได้

Accuracy	เท่ากับ 0.5495	หรือ 54.95%
Precision	เท่ากับ 0.5727	หรือ 57.27%
Recall	เท่ากับ 0.6768	หรือ 67.68%
F1-score	เท่ากับ 0.6204	หรือ 62.04%

4.3.2.2 โมเดลวิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best hyperparameter on Decision Tree Classifier

```
criterion = entropy
max_depth = 12
min_samples_leaf = 1
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5824 หรือ 58.24%
Precision	เท่ากับ 0.6322 หรือ 63.22%
Recall	เท่ากับ 0.5556 หรือ 55.56%
F1-score	เท่ากับ 0.5914 หรือ 59.14%

4.3.2.3 โมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network)

(1) The best hyperparameter on MLP Classifier

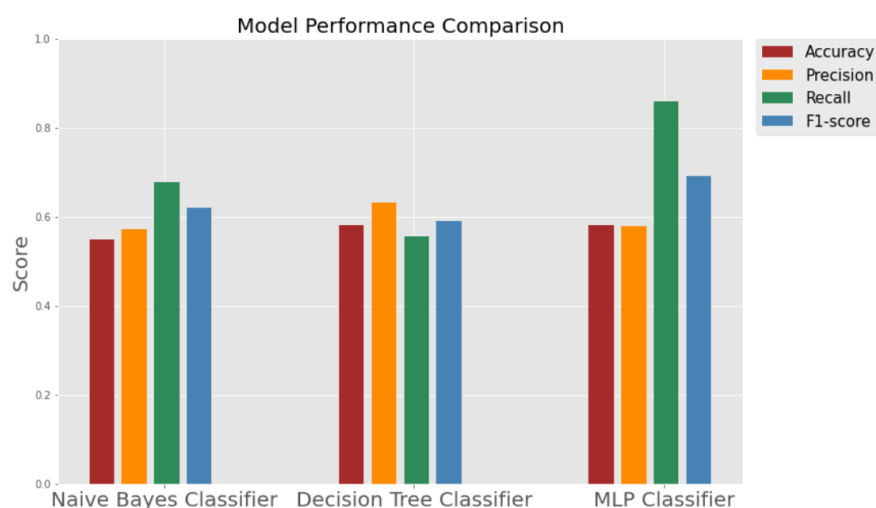
```
activation = tanh
hidden_layer_sizes = (20,)
alpha = 0.01
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5824 หรือ 58.24%
Precision	เท่ากับ 0.5782 หรือ 57.82%
Recall	เท่ากับ 0.8586 หรือ 85.86%
F1-score	เท่ากับ 0.6911 หรือ 69.11%

4.3.2.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

เปรียบเทียบโมเดลการทำนายแนวโน้มบนราคาหลักทรัพย์ KBANK.BK



ภาพที่ 4.27 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ KBANK.BK

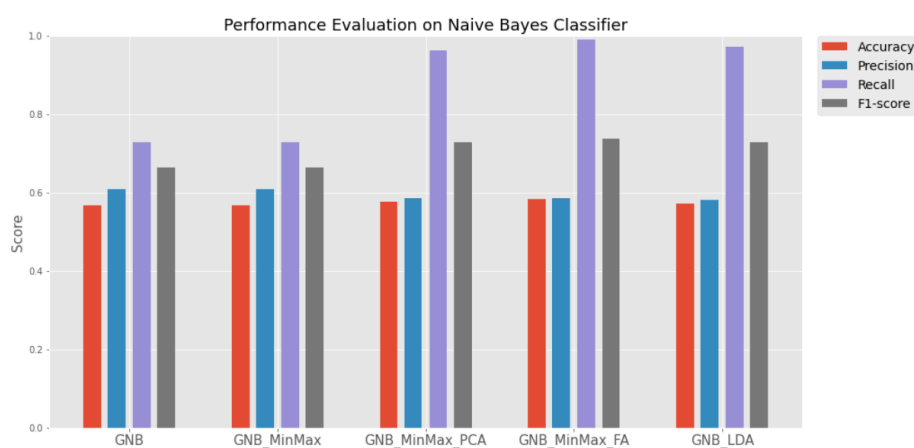
Predictive Modeling	Accuracy	Precision	Recall	F1-score
Naïve Bayes Classifier	0.5494	0.5727	0.6768	0.6204
Decision Tree Classifier	0.5824	0.6322	0.5556	0.5914
Multi-Layer Perceptron Classifier	0.5824	0.5782	0.8586	0.6911

ตารางที่ 4.11 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ KBANK.BK

จากผลการทดลองจากทั้งสามโมเดลบนชุดข้อมูลหลักทรัพย์ KBANK.BK จะเห็นว่าโมเดล Decision Tree Classifier และ MLP Classifier หลังการปรับค่าไฮเพอร์พารามิเตอร์ได้ให้ค่า Accuracy ได้สูงที่สุดเท่ากันที่ 0.5824 โดยในส่วนของค่า F1-score โมเดล MLP Classifier ให้ค่าสูงสุดคือ 0.6911

4.3.3 หลักทรัพย์ PTTEP.BK (กลุ่มทรัพยากร)

4.3.3.1 โมเดลวิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 4.28 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่ต่างกันในหลักทรัพย์ PTTEP.BK

(1) The best combination techniques on GNB Model :

Gaussian Naïve Bayes and MinMax-scaling and FA

(2) Evaluation Metrics

เมื่อรันผลผ่านโมเดล GNB_MinMax_FA จะได้

Accuracy	เท่ากับ 0.5824	หรือ 58.24%
Precision	เท่ากับ 0.5856	หรือ 58.56%
Recall	เท่ากับ 0.9907	หรือ 99.07%
F1-score	เท่ากับ 0.7361	หรือ 73.61%

4.3.3.2 โมเดลวิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best hyperparameter on Decision Tree

```
criterion = entropy
max_depth = 7
min_samples_leaf = 5
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5769 หรือ 57.69%
Precision	เท่ากับ 0.6087 หรือ 60.87%
Recall	เท่ากับ 0.7850 หรือ 78.50%
F1-score	เท่ากับ 0.6857 หรือ 68.57%

4.3.3.3 โมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network)

(1) The best hyperparameter on MLP Classifier

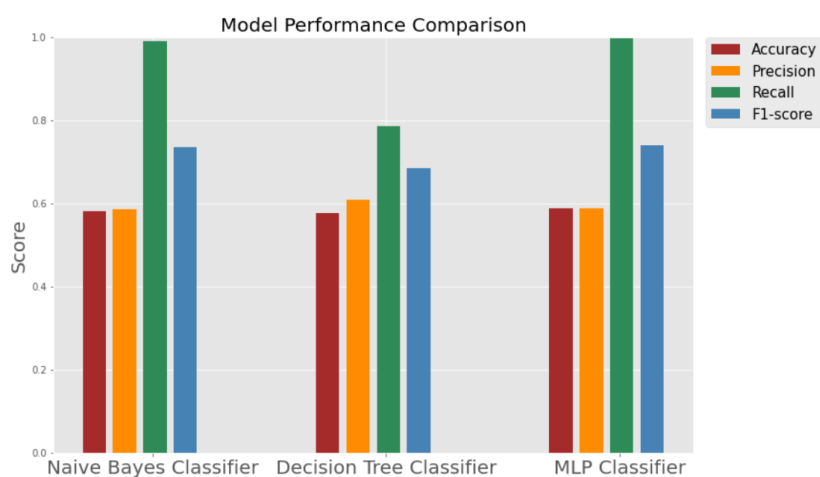
```
activation = identity
hidden_layer_sizes = (30,)
alpha = 0.0001
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5879 หรือ 58.79%
Precision	เท่ากับ 0.5879 หรือ 58.79%
Recall	เท่ากับ 1.0000 หรือ 100.00%
F1-score	เท่ากับ 0.7405 หรือ 74.05%

4.3.3.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

เปรียบเทียบโมเดลการทำนายแนวโน้มบนราคาหลักทรัพย์ PTTEP.BK



ภาพที่ 4.29 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ PTTEP.BK

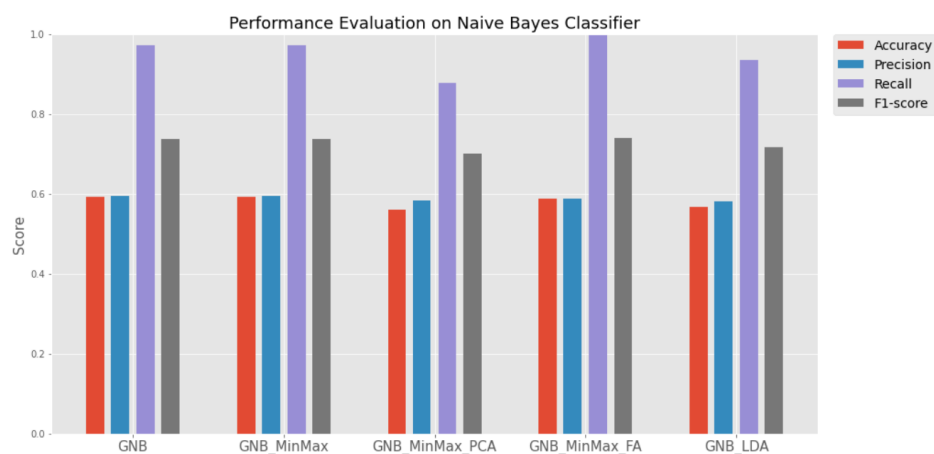
Predictive Modeling	Accuracy	Precision	Recall	F1-score
Naïve Bayes Classifier	0.5824	0.5856	0.9907	0.7361
Decision Tree Classifier	0.5769	0.6087	0.7850	0.6857
Multi-Layer Perceptron Classifier	0.5879	0.5879	1.0000	0.7405

ตารางที่ 4.12 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ PTTEP.BK

จากผลการทดลองจากทั้งสามโมเดลบนชุดข้อมูลหลักทรัพย์ PTTEP.BK จะเห็นว่าโมเดล MLP Classifier หลังจากที่ได้ปรับค่าไฮเพอร์พารามิเตอร์ ได้ให้ค่า Accuracy และค่า F1-score ได้สูงที่สุดคือ 0.5879 และ 0.7405 ตามลำดับ

4.3.4 หลักทรัพย์ CPALL.BK (กลุ่มบริการ)

4.3.4.1 โมเดลวิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 4.30 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่ต่างกันในหลักทรัพย์ CPALL.BK

(1) The best combination techniques on GNB Model :

Gaussian Naïve Bayes and MinMax-scaling

(2) Evaluation Metrics

เมื่อรันผลผ่านโมเดล GNB_MinMax จะได้

Accuracy	เท่ากับ 0.5934	หรือ 59.34%
Precision	เท่ากับ 0.5943	หรือ 59.43%
Recall	เท่ากับ 0.9720	หรือ 97.20%
F1-score	เท่ากับ 0.7376	หรือ 73.76%

4.3.4.2 โมเดลวิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best hyperparameter on Decision Tree

```
criterion = entropy
max_depth = 20
min_samples_leaf = 10
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5549 หรือ 55.49%
Precision	เท่ากับ 0.6300 หรือ 63.00%
Recall	เท่ากับ 0.5888 หรือ 58.88%
F1-score	เท่ากับ 0.6087 หรือ 60.87%

4.3.4.3 โมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network)

(1) The best hyperparameter on MLP Classifier

```
activation = relu
hidden_layer_sizes = (40,)
alpha = 0.0001
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.6154 หรือ 61.54%
Precision	เท่ากับ 0.6045 หรือ 60.45%
Recall	เท่ากับ 1.0000 หรือ 100.00%
F1-score	เท่ากับ 0.7535 หรือ 75.35%

4.3.4.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

เปรียบเทียบโมเดลการทำนายแนวโน้มบนราคาหลักทรัพย์ CPALL.BK



ภาพที่ 4.31 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ CPALL.BK

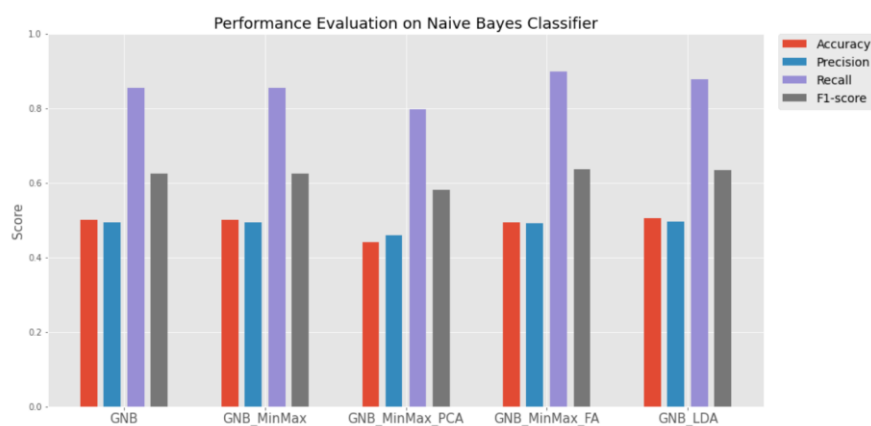
Predictive Modeling	Accuracy	Precision	Recall	F1-score
Naïve Bayes Classifier	0.5934	0.5943	0.9719	0.7376
Decision Tree Classifier	0.5549	0.6300	0.5888	0.6087
Multi-Layer Perceptron Classifier	0.6154	0.6045	1.0000	0.7535

ตารางที่ 4.13 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ CPALL.BK

จากผลการทดลองจากทั้งสามโมเดลบนชุดข้อมูลหลักทรัพย์ CPALL.BK จะเห็นว่าโมเดล MLP Classifier หลังจากที่ได้ปรับค่าไฮเพอร์พารามิเตอร์ ได้ให้ค่า Accuracy และค่า F1-score ได้สูงที่สุดคือ 0.6154 และ 0.7535 ตามลำดับ

4.3.5 หลักทรัพย์ DELTA.BK (กลุ่มเทคโนโลยี)

4.3.5.1 โมเดลวิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 4.32 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่ต่างกันบนหลักทรัพย์ DELTA.BK

(1) The best combination techniques on GNB Model :

Gaussian Naïve Bayes and LDA

(2) Evaluation Metrics

เมื่อรันผลผ่านโมเดล GNB_LDA จะได้

Accuracy	เท่ากับ 0.5055	หรือ 50.55%
Precision	เท่ากับ 0.4968	หรือ 49.68%
Recall	เท่ากับ 0.8764	หรือ 87.64%
F1-score	เท่ากับ 0.6341	หรือ 63.41%

4.3.5.2 โมเดลวิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best hyperparameter on Decision Tree Classifier

```
criterion = entropy
max_depth = 12
min_samples_leaf = 7
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5495 หรือ 54.95%
Precision	เท่ากับ 0.5289 หรือ 52.89%
Recall	เท่ากับ 0.7191 หรือ 71.91%
F1-score	เท่ากับ 0.6095 หรือ 60.95%

4.3.5.3 โมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network)

(1) The best hyperparameter on MLP Classifier

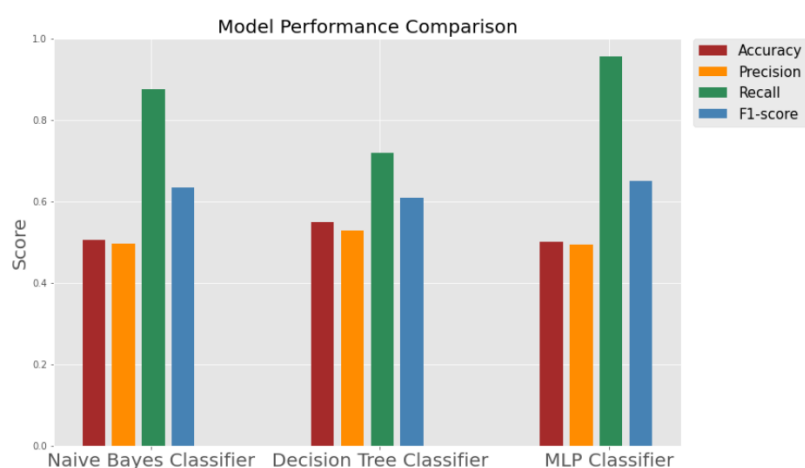
```
activation = identity
hidden_layer_sizes = (20,)
alpha = 0.0001
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5000 หรือ 50.00%
Precision	เท่ากับ 0.4942 หรือ 49.42%
Recall	เท่ากับ 0.9551 หรือ 95.51%
F1-score	เท่ากับ 0.6513 หรือ 65.13%

4.3.5.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

เปรียบเทียบโมเดลการทำนายแนวโน้มบนราคาหลักทรัพย์ DELTA.BK



ภาพที่ 4.33 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ DELTA.BK

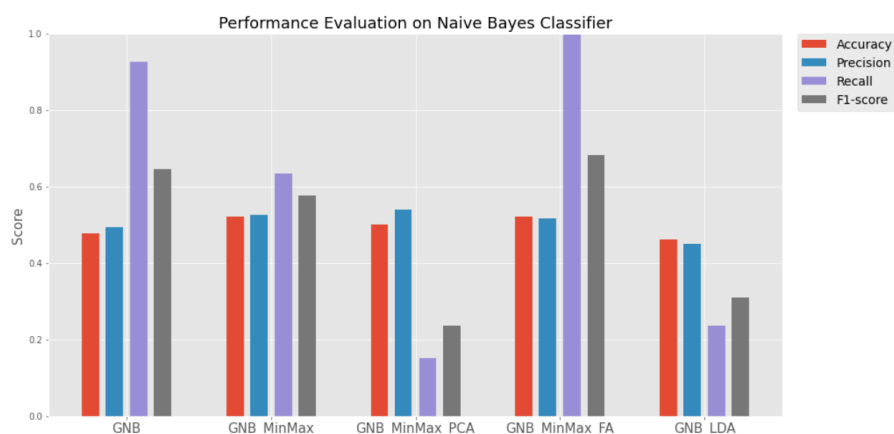
Predictive Modeling	Accuracy	Precision	Recall	F1-score
Naïve Bayes Classifier	0.5055	0.4968	0.8764	0.6341
Decision Tree Classifier	0.5495	0.5289	0.7191	0.6095
Multi-Layer Perceptron Classifier	0.5000	0.4942	0.9551	0.6513

ตารางที่ 4.14 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ DELTA.BK

จากผลการทดลองจากทั้งสามโมเดลบนชุดข้อมูลหลักทรัพย์ DELTA.BK จะเห็นว่าโมเดล Decision Tree Classifier หลังจากที่ได้ปรับค่าไฮเพอร์พารามิเตอร์ ได้ให้ค่า Accuracy สูงที่สุดคือ 0.5495 ในขณะที่โมเดล MLP Classifier หลังจากที่ได้ปรับค่าไฮเพอร์พารามิเตอร์ ได้ให้ค่า F1-score สูงที่สุด คือ 0.6513

4.3.6 หลักทรัพย์ SCC.BK (กลุ่มอสังหาริมทรัพย์และการก่อสร้าง)

4.3.6.1 โมเดลวิธีจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)



ภาพที่ 4.34 กราฟแสดงผลการวัดประสิทธิภาพของวิธีจำแนกแบบเบย์อย่างง่ายควบคู่กับเทคนิควิธีที่แตกต่างกันบนหลักทรัพย์ SCC.BK

(1) The best combination techniques on GNB Model :

Gaussian Naïve Bayes and MinMax-scaling

(2) Evaluation Metrics

เมื่อรันผลผ่านโมเดล GNB_MinMax จะได้

Accuracy	เท่ากับ 0.5219 หรือ 52.19%
Precision	เท่ากับ 0.5268 หรือ 52.68%
Recall	เท่ากับ 0.6344 หรือ 63.44%
F1-score	เท่ากับ 0.5756 หรือ 57.56%

4.3.6.2 โมเดลวิธีจำแนกต้นไม้การตัดสินใจ (Decision Tree Classifier)

(1) The best hyperparameter on Decision Tree

```
criterion = gini
max_depth = 12
min_samples_leaf = 9
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5769 หรือ 57.69%
Precision	เท่ากับ 0.5976 หรือ 59.76%
Recall	เท่ากับ 0.5269 หรือ 52.69%
F1-score	เท่ากับ 0.5600 หรือ 56.00%

4.3.6.3 โมเดลวิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network)

(1) The best hyperparameter on MLP Classifier

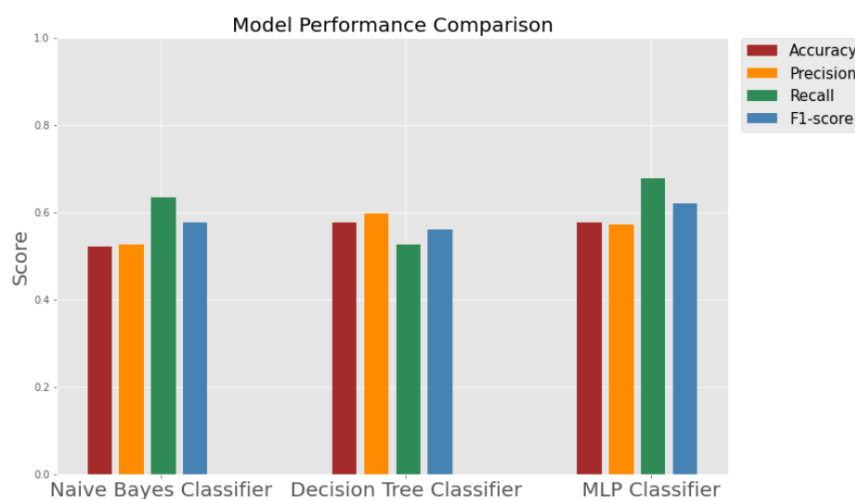
```
activation = relu
hidden_layer_sizes = (40,)
alpha = 0.0001
```

(2) Evaluation Metrics

Accuracy	เท่ากับ 0.5769 หรือ 57.69%
Precision	เท่ากับ 0.5727 หรือ 57.27%
Recall	เท่ากับ 0.6774 หรือ 67.74%
F1-score	เท่ากับ 0.6207 หรือ 62.07%

4.3.6.4 เปรียบเทียบประสิทธิภาพ (Performance Evaluation)

เปรียบเทียบโมเดลการทำนายแนวโน้มบนราคาหลักทรัพย์ SCC.BK



ภาพที่ 4.35 กราฟแสดงผลการวัดประสิทธิภาพของวิธีการทำนายแนวโน้มบนหลักทรัพย์ SCC.BK

Predictive Modeling	Accuracy	Precision	Recall	F1-score
Naïve Bayes Classifier	0.5220	0.5268	0.6344	0.5756
Decision Tree Classifier	0.5769	0.5976	0.5269	0.5600
Multi-Layer Perceptron Classifier	0.5769	0.5727	0.6774	0.6207

ตารางที่ 4.15 ผลคะแนนการวัดประสิทธิภาพของวิธีการทำนายผลบนหลักทรัพย์ SCC.BK

จากผลการทดลองจากทั้งสามโมเดลบนชุดข้อมูลหลักทรัพย์ SCC.BK จะเห็นว่าโมเดล Decision Tree Classifier และ MLP Classifier หลังการปรับค่าไฮเพอร์พารามิเตอร์ได้ให้ค่า Accuracy ได้สูงที่สุดเท่ากันที่ 0.5769 โดยในส่วนของค่า F1-score โมเดล MLP Classifier ให้ค่าสูงสุดคือ 0.6207

4.4 การศึกษาประสิทธิภาพและเปรียบเทียบ (Model Evaluation and Comparison)

4.4.1 ประสิทธิภาพวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูล

Statistical Analysis				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.5659	0.6000	0.6061	0.6030
PTTEP	0.5824	0.6220	0.7383	0.6752
CPALL	0.5824	0.6084	0.8131	0.6960
DELTA	0.6099	0.5818	0.7191	0.6432
SCC	0.5495	0.5567	0.5806	0.5684
Mean	0.57802	0.59378	0.69144	0.63716

ตารางที่ 4.16 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีการวิเคราะห์ทางสถิติ

Naïve Bayes Classifier				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.6593	0.7079	0.6364	0.6702
PTTEP	0.5934	0.6387	0.7103	0.6725
CPALL	0.5934	0.6634	0.6262	0.6442
DELTA	0.5934	0.5773	0.6292	0.6021
SCC	0.5714	0.5862	0.5484	0.5667
Mean	0.60218	0.63470	0.63010	0.63114

ตารางที่ 4.17 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกแบบเบย์อย่างง่าย

Decision Tree Classifier				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.6429	0.6700	0.6768	0.6734
PTTEP	0.5494	0.6190	0.6075	0.5854
CPALL	0.5604	0.6216	0.6449	0.6330
DELTA	0.5769	0.5612	0.6180	0.5882
SCC	0.5604	0.5684	0.5806	0.5745
Mean	0.57800	0.60804	0.62556	0.61090

ตารางที่ 4.18 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกต้นไม้การตัดสินใจ

4.4.2 ประสิทธิภาพวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐาน

Naïve Bayes Classifier				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.5495	0.5727	0.6768	0.6204
PTTEP	0.5824	0.5856	0.9907	0.7361
CPALL	0.5934	0.5943	0.9719	0.7376
DELTA	0.5055	0.4968	0.8764	0.6341
SCC	0.5220	0.5268	0.6344	0.5756
Mean	0.55056	0.55524	0.83004	0.66076

ตารางที่ 4.19 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีการวิเคราะห์ทางสถิติ

Decision Tree Classifier				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.5824	0.6322	0.5556	0.5914
PTTEP	0.5769	0.6087	0.7850	0.6857
CPALL	0.5549	0.6300	0.5888	0.6087
DELTA	0.5495	0.5289	0.7191	0.6095
SCC	0.5769	0.5976	0.5269	0.5600
Mean	0.56812	0.59948	0.63508	0.61106

ตารางที่ 4.20 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกแบบเบย์อย่างง่าย

Multi-Layer Perceptron Neural Network				
Dataset	Accuracy	Precision	Recall	F1-score
KBANK	0.5824	0.5782	0.8586	0.6911
PTTEP	0.5879	0.5879	1.0000	0.7405
CPALL	0.6154	0.6045	1.0000	0.7535
DELTA	0.5000	0.4942	0.9551	0.6513
SCC	0.5769	0.5727	0.6774	0.6207
Mean	0.57252	0.5675	0.89822	0.69142

ตารางที่ 4.21 ผลคะแนนการวัดประสิทธิภาพการทำนายแนวโน้มของวิธีจำแนกโครงข่ายประสาทเทียม
เพอร์เซพตรอนแบบหลายชั้น

4.5 สรุปผลการเปรียบเทียบ

จากการผลเปรียบเทียบจะเห็นได้ว่าวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูลแสดงประสิทธิภาพความแม่นยำได้สูงขึ้นเมื่อเทียบกับวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐาน โดยพิจารณาภาพรวมของค่า Accuracy โดยเฉลี่ย (Mean) ในหลักทรัพย์แต่ละรายการระหว่างสองเทคนิควิธี และเมื่อพิจารณาจากค่า Accuracy ในหลักทรัพย์แบบรายตัวจะเห็นได้ว่าวิธีจำแนกแบบเบย์อย่างง่ายสามารถทำนายแนวโน้มได้แม่นยำที่สุดถึง 65.93% บนราคาหลักทรัพย์ KBANK.BK ในส่วนของเทคนิควิธีการพื้นฐานนั้น วิธีจำแนกโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้นเป็นวิธีการทำนายแนวโน้มที่แม่นยำที่สุดอยู่ที่ 61.54% บนราคาหลักทรัพย์ CPALL.BK

ในส่วนของการเปรียบเทียบผลด้วยค่า Precision ระหว่างสองเทคนิควิธีจะเห็นได้ว่าเทคนิควิธีการทำเหมืองข้อมูลแสดงประสิทธิภาพโดยเฉลี่ยได้สูงกว่าเช่นเดียวกัน โดยเมื่อพิจารณาค่า Precision ในหลักทรัพย์แบบรายตัวจะเห็นได้ว่าวิธีจำแนกแบบเบย์อย่างง่ายทำนายแนวโน้มได้แม่นยำสูงถึง 70.79% จากการทำนายในคลาสแนวโน้มราคาขึ้นบนหลักทรัพย์ KBANK.BK

สำหรับการเปรียบเทียบผลด้วยค่า F1-score จะเห็นได้ชัดว่าเทคนิควิธีการพื้นฐานแสดงประสิทธิภาพโดยรวมได้ดีกว่า เนื่องจากค่า Recall ที่คำนวณได้สูงสุดถึง 100% บนการทำนายแนวโน้มหลักทรัพย์บางรายการ โดยเกิดขึ้นจากการที่โมเดลทำนายผลลัพธ์ในเชิงบวก (แนวโน้มขึ้น) เป็นจำนวนมากต่อเนื่องกันหลังจากที่โมเดลมีการปรับค่าไฮเปอร์พารามิเตอร์ที่ให้ค่าความแม่นยำสูงสุด ซึ่งอาจจะเป็นผลมาจากสัดส่วนของคลาสแนวโน้มที่ไม่เท่ากันบนหลักทรัพย์บางกลุ่ม

4.6 ผลการวิเคราะห์การทำนายแนวโน้มแบบรายวันผ่านแพลตฟอร์มซื้อขายหลักทรัพย์จำลอง

(1) Analyzing on KBANK.BK

Date	Open	High	Low	Close	Probability of Uptrend in next day	Average of Price Change	Expectation of Random variable (Trend)	Variance of Random variable (Trend)	Naïve Bayes Classifier (Trend)	Decision Tree Classifier (Trend)
14/11/2022	146.00	147.00	144.00	144.50	0.7333	+0.62	0.4082	0.8333	-1	-1
15/11/2022	144.50	146.00	144.50	144.50	0.6000	+0.48	0.1784	0.9682	-1	-1
16/11/2022	144.50	145.00	144.00	145.00	0.6667	+1.05	0.2473	0.9388	+1	+1
17/11/2022	145.00	145.00	143.00	143.50	0.4667	+0.68	-0.0768	0.9941	-1	+1
18/11/2022	143.50	143.50	142.00	143.00	0.2000	-1.93	-0.5718	0.6731	-1	-1
21/11/2022	143.00	143.50	142.00	142.50	0.4000	-0.47	-0.1652	0.9727	-1	-1
22/11/2022	142.50	144.00	142.50	143.50	0.4667	-0.08	-0.0747	0.9944	-1	-1
23/11/2022	143.50	143.50	142.50	143.00	0.4000	-0.43	-0.1069	0.9886	-1	-1
24/11/2022	143.50	144.00	143.00	143.50	0.4667	-0.62	-0.0436	0.9981	-1	+1
25/11/2022	143.50	144.50	143.00	143.50	0.4667	-0.35	-0.1505	0.9774	-1	+1
28/11/2022	142.50	143.50	141.50	143.50	0.3333	-1.28	-0.2778	0.9228	+1	+1
29/11/2022	143.50	146.00	143.00	145.50	0.4667	-0.08	-0.0863	0.9925	-1	-1
30/11/2022	145.00	145.00	144.00	144.50	0.7333	+1.35	0.4584	0.7898	-1	+1
01/12/2022	144.50	145.00	144.00	144.00	0.4667	-0.27	-0.0402	0.9984	-1	+1
02/12/2022	143.50	144.00	142.50	142.50	0.6667	+0.55	0.2987	0.9108	-1	+1
06/12/2022	141.50	142.00	141.00	142.00	0.6667	+1.07	0.4016	0.8387	+1	+1
07/12/2022	142.00	144.00	142.00	142.50	0.5333	+0.67	0.0699	0.9951	-1	+1
08/12/2022	142.50	143.00	142.00	142.00	0.6000	+0.33	0.2953	0.9128	+1	+1
09/12/2022	143.00	143.00	142.00	142.50						

ตารางที่ 4.22 ผลการวิเคราะห์แนวโน้มราคาหลักทรัพย์ KBANK.BK แบบรายวัน

(2) Analyzing on DELTA.BK

Date	Open	High	Low	Close	Probability of Uptrend in next day	Average of Price Change	Expectation of Random variable (Trend)	Variance of Random variable (Trend)	Naïve Bayes Classifier	Decision Tree Classifier
14/11/2022	614.00	630.00	604.00	624.00	0.4400	+2.88	-0.2163	0.9532	-1	+1
15/11/2022	628.00	636.00	624.00	632.00	0.4800	+10.41	0.0018	0.9999	-1	-1
16/11/2022	628.00	640.00	624.00	634.00	0.2000	-1.18	-0.6248	0.6097	-1	-1
17/11/2022	638.00	640.00	630.00	638.00	0.3600	-5.96	-0.2925	0.9144	+1	+1
18/11/2022	638.00	668.00	638.00	660.00	0.5200	+3.28	-0.0115	0.9998	-1	+1
21/11/2022	654.00	656.00	636.00	640.00	0.5200	+5.18	-0.0212	0.9995	-1	-1
22/11/2022	646.00	662.00	644.00	652.00	0.6000	+3.01	0.2103	0.9558	-1	-1
23/11/2022	660.00	670.00	650.00	656.00	0.4400	-3.91	-0.1259	0.9842	-1	+1
24/11/2022	662.00	668.00	656.00	666.00	0.2800	-3.35	-0.5259	0.7233	+1	+1
25/11/2022	666.00	666.00	656.00	662.00	0.6400	+2.35	0.3466	0.8798	-1	-1
28/11/2022	664.00	676.00	656.00	666.00	0.2000	-4.92	-0.6413	0.5886	+1	-1
29/11/2022	668.00	668.00	644.00	656.00	0.5600	-0.99	0.1204	0.9855	+1	-1
30/11/2022	658.00	672.00	658.00	672.00	0.6400	+1.21	0.3205	0.8973	+1	+1
01/12/2022	684.00	782.00	682.00	764.00	0.7200	+12.08	0.3694	0.8636	+1	-1
02/12/2022	776.00	796.00	760.00	772.00	0.6400	+11.70	0.1976	0.9609	-1	-1
06/12/2022	764.00	778.00	738.00	740.00	0.6400	-3.27	0.3076	0.9054	+1	+1
07/12/2022	736.00	742.00	686.00	694.00	0.5600	+1.94	0.1294	0.9832	-1	-1
08/12/2022	704.00	714.00	678.00	682.00	0.6800	+2.46	0.3550	0.8740	+1	+1
09/12/2022	688.00	720.00	678.00	702.00						

ตารางที่ 4.23 ผลการวิเคราะห์แนวโน้มราคาหลักทรัพย์ DELTA.BK แบบรายวัน

บทที่ 5

สรุป

5.1 สรุปผลการทดลอง

โครงการงานศึกษาเทคนิควิธีการทำเหมืองข้อมูลสำหรับค้นหาแนวโน้มราคาหลักทรัพย์เพื่อทำนายผลโดยใช้กรอบความน่าจะเป็นได้ใช้เทคนิคการทำเหมืองข้อมูลโดยใช้วิธีค้นพบโมทีฟในอนุกรมเวลาผ่านอัลกอริทึมที่ใช้สำหรับการค้นหาลำดับย่อยในข้อมูลอนุกรมเวลาที่มีลักษณะคล้ายกันผ่านการคำนวณระยะห่างความคล้ายคลึงกัน (Similarity Distance Measure) เพื่อค้นหารูปแบบในอดีตที่คล้ายคลึงกับปัจจุบันบนชุดข้อมูลตลาดหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย (SET) โดยนำรูปแบบและข้อมูลที่ได้ทั้งหมดมาทำการคำนวณผ่านกรอบวิธีความน่าจะเป็น ได้แก่ วิธีการวิเคราะห์ทางสถิติ และวิธีการเรียนรู้ของเครื่อง ได้แก่ วิธีจำแนกแบบเบย์อย่างง่าย และวิธีจำแนกต้นไม้การตัดสินใจ เพื่อค้นหารูปแบบและทำนายโอกาสในการเปลี่ยนแปลงของแนวโน้มราคาหลักทรัพย์รายวัน (Daily Trends Movement of Stock data) ซึ่งในโครงการงานศึกษาได้มีการศึกษาเปรียบเทียบกับเทคนิควิธีการพื้นฐานที่ดึงคุณลักษณะโดยใช้ตัวชี้วัดทางเทคนิค (Technical Indicators) ของหลักทรัพย์สำหรับทำนายผลผ่านอัลกอริทึมทางปัญญาประดิษฐ์ ได้แก่ วิธีการเรียนรู้ของเครื่อง และวิธีการเรียนรู้เชิงลึก เพื่อศึกษาและเปรียบเทียบประสิทธิภาพความแม่นยำของเทคนิควิธีการต่างๆ ในการสร้างวิธีการทำนายแนวโน้มราคาหลักทรัพย์แบบรายวัน

ในการศึกษานี้ ผู้ทดลองได้ใช้ชุดข้อมูลหุ้นจากตลาดหลักทรัพย์แห่งประเทศไทยในกลุ่มอุตสาหกรรมที่แตกต่างกัน โดยได้ทำการเก็บรวบรวมข้อมูลย้อนหลังในทุกรายการเป็นเวลา 5 ปี นับตั้งแต่วันที่ 10 พฤศจิกายน 2560 จนถึงวันที่ 10 พฤศจิกายน 2565 จากนั้นทำการทดลองเพื่อค้นหารูปแบบของราคาหุ้นในอดีตผ่านวิธีการทำเหมืองข้อมูลด้วยเทคนิควิธีการค้นพบโมทีฟในข้อมูลอนุกรมเวลาโดยใช้อัลกอริทึมสำหรับการคำนวณระยะทางระหว่างลำดับย่อยของอนุกรมเวลาในการค้นหารูปแบบที่คล้ายกับรูปแบบปัจจุบันทั้งหมด เพื่อนำรูปแบบและข้อมูลที่ได้มาวิเคราะห์และสร้างโมเดลการทำนาย โดยในการศึกษานี้ได้มีการทดลองด้วยการปรับค่าพารามิเตอร์ในค่าที่แตกต่างกัน ได้แก่ จำนวนของรูปแบบราคา (Number of patterns) และขนาดของรูปแบบราคา (Window size) เพื่อค้นหาวิธีการทำนายผลที่มีการ

กำหนดค่าพารามิเตอร์แล้วให้ค่าความแม่นยำในการทำนายผลได้ดีที่สุดบนข้อมูลตลาดหุ้นแต่ละรายการ สำหรับในส่วนของเทคนิควิธีการพื้นฐานที่ดึงคุณลักษณะโดยใช้ตัวชี้วัดทางเทคนิค จะทำการคำนวณตัวชี้วัดทางเทคนิคเพื่อใช้เป็นตัวแปรนำเข้าสำหรับการสร้างโมเดลทำนายผล โดยได้มีการปรับจูนค่าไฮเพอร์พารามิเตอร์ (Hyperparameter tuning) และใช้เทคนิคการดึงคุณลักษณะ (Feature extraction) ที่แตกต่างกันบนโมเดลที่จะทำการเปรียบเทียบเพื่อให้ได้โมเดลที่มีประสิทธิภาพที่ดีที่สุดสำหรับใช้ศึกษาเปรียบเทียบ

จากผลการทดลองพบว่า วิธีการทำนายแนวโน้มราคาหลักทรัพย์ผ่านเทคนิควิธีการทำเหมืองข้อมูลมีค่า Mean Accuracy สูงกว่าเทคนิควิธีการพื้นฐานโดยเฉลี่ยอยู่ที่ 1-3% โดยสำหรับเทคนิควิธีการทำเหมืองข้อมูล วิธีการจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier) เป็นวิธีที่ให้ค่า Accuracy เฉลี่ยสูงที่สุด คือ 0.6022 และในส่วนของเทคนิควิธีการพื้นฐาน วิธีจำแนกโครงข่ายเพอร์เซพตรอนแบบหลายชั้น (MLP Classifier) เป็นวิธีที่ให้ค่า Mean Accuracy เฉลี่ยสูงที่สุด คือ 0.5725 อย่างไรก็ตามเมื่อพิจารณาค่า F1-score เทคนิควิธีการพื้นฐานนั้นกลับให้ค่าเฉลี่ยโดยรวมที่สูงกว่า อันเนื่องมาจากค่า Recall ที่ค่อนข้างสูงถึง 1.0000 จากการทำนายแนวโน้มขึ้นเป็นส่วนใหญ่ของโมเดลบนหลักทรัพย์บางกลุ่ม แต่เมื่อพิจารณาค่า Precision ที่ใช้วัดความแม่นยำในการทำนายผลเชิงบวก (แนวโน้มขึ้น) ซึ่งมีความสำคัญในการวัดผลตอบแทนและกำไร จะพบว่าการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูลมีค่า Mean Precision โดยเฉลี่ยสูงกว่าเทคนิควิธีการพื้นฐาน 1-8% ซึ่งจากผลข้างต้น เมื่อพิจารณาโดยรวมจะเห็นได้ว่าวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการทำเหมืองข้อมูลที่ได้ทดลองศึกษาแสดงค่าความถูกต้องและค่าความแม่นยำได้สูงขึ้น เมื่อเทียบกับวิธีการทำนายแนวโน้มผ่านเทคนิควิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค

ผู้จัดทำคาดหวังว่าโครงการการศึกษาเปรียบเทียบวิธีการทำนายแนวโน้มราคาหลักทรัพย์ผ่านเทคนิควิธีการทำเหมืองข้อมูลด้วยวิธีค้นพบโมทีฟในอนุกรมเวลา และเทคนิควิธีการพื้นฐานโดยใช้ตัวชี้วัดทางเทคนิค จะสามารถแสดงให้เห็นถึงความแตกต่างของเทคนิควิธีการเตรียมข้อมูล วิธีการดึงคุณลักษณะ วิธีการคัดเลือกตัวแปร และประสิทธิภาพในการทำนายผลบนวิธีการที่แตกต่างกัน เพื่อเป็นแนวทางสำหรับผู้สนใจศึกษา ทั้งในด้านการศึกษาพัฒนาวิธีการทำนายผลหรืออัลกอริทึมในการทำเหมืองข้อมูล และด้านการวิเคราะห์และ

ตัดสินใจลงทุนในหลักทรัพย์แบบรายวัน ซึ่งสามารถนำไปประยุกต์และพัฒนาต่อยอดได้หลายแขนงอีกในอนาคต

5.2 แนวคิดต่อยอดและการปรับปรุงในอนาคต

สำหรับแนวคิดในการต่อยอดในอนาคต สามารถศึกษาพัฒนาเพิ่มได้จากเดิมที่มีการใช้ข้อมูลกราฟราคาปิดรายวันซึ่งเป็นข้อมูลอนุกรมเวลาแบบตัวแปรเดียว (Univariate time-series) เปลี่ยนเป็นการใช้ข้อมูลกราฟแบบแท่งเทียนที่ประกอบด้วยข้อมูล ราคาเปิด ราคาปิด ราคาสูงสุด และราคาต่ำสุด รวมถึงข้อมูลปริมาณซื้อขายและอินดิเคเตอร์ที่เกี่ยวข้อง ซึ่งเป็นข้อมูลอนุกรมเวลาแบบหลายตัวแปร (Multivariate time-series) รวมถึงยังสามารถปรับปรุงในส่วนของการทำนายแนวโน้มแบบรายวัน (Daily trends) พัฒนาเป็นการทำนายแนวโน้มในระยะสั้นหรือระยะยาวได้ เช่น การทำนายแนวโน้มในช่วงวันที่กว้างขึ้น หรือช่วงสัปดาห์ ร่วมกับการใช้ตัวแปรปัจจัยทางพื้นฐานของข้อมูลตลาดหุ้นมาใช้ในการสร้างโมเดล ซึ่งเป็นสิ่งที่คิดว่าท้าทายและน่าสนใจมากขึ้น นอกจากนี้ยังสามารถพัฒนาและปรับปรุงเทคนิควิธีในการทำนายผ่านโมเดลที่มีความซับซ้อนมากยิ่งขึ้นได้ ซึ่งอาจจะทำให้ประสิทธิภาพความแม่นยำเพิ่มสูงขึ้นได้เช่นเดียวกัน

แหล่งอ้างอิง

- [1] SETInvestnow. (2020). การวิเคราะห์หลักทรัพย์โดยใช้ปัจจัยพื้นฐาน, สืบค้นเมื่อ 16 พฤษภาคม 2565. จาก <https://www.setinvestnow.com/th/glossary/fundamental-analysis>
- [2] Admiral Markets. (2022). สรุปภาพรวม : การวิเคราะห์ทางเทคนิค (Technical Analysis) คืออะไร?, สืบค้นเมื่อ 16 พฤษภาคม 2565. จาก <https://admiralmarkets.sc/th/education/articles/forex-analysis/technical-analysis-forex>
- [3] ภาคิน รัตนกร. สรุปอินดิเคเตอร์ Indicator. กรุงเทพมหานคร, 2565
- [4] Chutipas Borsub. (2021). ประเภทของ Indicator, สืบค้นเมื่อ 6 พฤศจิกายน 2565. จาก <https://bitkubacademy.com/th/blog/types-of-indicators>
- [5] Sirawich vounghchuy. (2018). Machine Learning : 01 — Data Preprocessing (python coding[basic]), สืบค้นเมื่อ 18 พฤษภาคม 2565. จาก <https://sirawit0676.medium.com/machine-learning-01-data-preprocessing-python-coding-basic-687aee03c478>.
- [6] Surapong Kanoktipsatharporn. (2019). Normalization คืออะไร ปรับช่วงข้อมูล Feature Scaling ด้วยวิธี Normalization, Standardization ก่อนเทรน Machine Learning – Preprocessing ep.2. สืบค้นเมื่อ 18 พฤษภาคม 2565. จาก <https://www.bualabs.com/archives/2100/what-is-normalization-feature-scaling-rescaling-normalization-standardization-feedforward-train-machine-learning-preprocessing-ep-2/>.
- [7] Mary K. Pratt. (2022). data transformation. สืบค้นเมื่อ 8 พฤศจิกายน 2565. จาก <https://www.techtarget.com/searchdatamanagement/definition/data-transformation>

- [8] ACM SIGKDD. (2006). Data Mining Curriculum. สืบค้นเมื่อ 6 พฤศจิกายน 2565.
- [9] Han, Kamber, Pei, Jaiwei, Micheline, Jian. (2011). Data Mining: Concepts and Techniques. สืบค้นเมื่อ 6 พฤศจิกายน 2565.
- [10] Abdullah Mueen, Eamonn Keogh, Qiang Zhu, Sydney Cash, Brandon Westover. (2009). Exact Discovery of Time Series Motifs.
- [11] Eamonn Keogh, Jessica Lin, Ada Fu. (2005). HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications.
- [12] Nat Kulvanich, Akarin Phaibulpanich. (2018). Comparison of Distance Measures in Cluster Analysis for Time-Series Data.
- [13] Kei Nakagawa, Mitsuyoshi Imamura, Kenichi Yoshida. (2019). Stock price prediction using k-medoids clustering with indexing dynamic time warping.
- [14] George Lawton et al. (2022). predictive modeling. สืบค้นเมื่อ 6 พฤศจิกายน 2565. จาก <https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling>
- [15] TechTarget Contributor. (2020). Statistical analysis. สืบค้นเมื่อ 6 พฤศจิกายน 2565. จาก <https://www.techtarget.com/whatis/definition/statistical-analysis>
- [16] บัญชา ปะสีละเตสัง. สร้างการเรียนรู้สำหรับ AI ด้วย Python Machine Learning. กรุงเทพมหานคร.
- [17] DeepAI. Evaluation Metrics. สืบค้นเมื่อ 6 พฤศจิกายน 2565. จาก <https://deepai.org/machine-learning-glossary-and-terms/evaluation-metrics>
- [18] Pagon Gatchalee. (2019). Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย ในMachine learning. สืบค้นเมื่อ 6 พฤศจิกายน 2565. จาก <https://medium.com/@pagongatchalee/confusion-matrix-เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย-ในmachine-learning-fba6e3f9508c>
- [19] FERNANDO G. D. C. FERREIRA, AMIR H. GANDOMI, RODRIGO T. N. CARDOSO. (2021). Artificial Intelligence Applied to Stock Market Trading: A Review.

- [20] Seungwoo Jeon, Bonghee Hong, Victor Chang. (2018). Pattern graph tracking-based stock price prediction using big data
- [21] Lior Sidi. (2020). Improving S&P stock prediction with time series stock similarity.
- [22] Min Wen, Ping Li, Lingfei Zhang, Yan Chen. (2019). Stock Market Trend Prediction Using High-Order Information of Time Series.
- [23] Xinjie Di. (2014). Stock Trend Prediction with Technical Indicator using SVM.
- [24] Ernest Kwame Ampomah, Gabriel Nyame, Zhiguang Qin, Prince Clement Addo, Enoch Opanin Gyamf, Micheal Gyan. (2021). Stock Market Prediction with Gaussian Naive Bayes Machine Learning Algorithm.
- [25] Rupesh A. Kamble. (2017). Short and Long Term Stock Trend Prediction using Decision Tree.
- [26] Sarayut Nonsiri, PhD. Computer Scientist. ภาษาโปรแกรม Python คืออะไร?. สืบค้นเมื่อ 14 พฤษภาคม 2565. จาก <https://www.9experttraining.com/articles/python-คืออะไร>
- [27] Numpy. NumPy documentation. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://numpy.org/doc/stable/>
- [28] Pandas. Package overview. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก https://pandas.pydata.org/docs/getting_started/overview.html
- [29] ActiveState. (2022). What Is Matplotlib In Python?. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://www.activestate.com/resources/quick-reads/what-is-matplotlib-in-python-how-to-use-it-for-plotting/>
- [30] Seaborn. seaborn: statistical data visualization. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://seaborn.pydata.org/>
- [31] Tutorials Point. (2019). Scikit Learn – Introduction. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก https://www.tutorialspoint.com/scikit_learn/scikit_learn_introduction.htm
- [32] Leo Smigel. (2022). yfinance Python Tutorial. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://analyzingalpha.com/yfinance-python>

- [33] Wannes Meert. Time Series Distances. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://pypi.org/project/dtaidistance/>
- [34] stackpython. (2020). สอนการติดตั้ง Jupyter Notebook. สืบค้นเมื่อ 9 พฤศจิกายน 2565. จาก <https://stackpython.medium.com/สอนการติดตั้ง-jupyter-notebook-1850bf72d8b1>
- [35] Pan Park. (2018). สาย Machine Learning ไม่ควรพลาดกับ Google Colab. สืบค้นเมื่อ 15 พฤษภาคม 2565. จาก <https://medium.com/@gnothaigamero/สาย-ml-ไม่ควรพลาดกับ-google-colab-a9f826cbf156>