

การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้ม
ของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย
**Predict stock price trends in Stock Exchange of Thailand using
Ensemble Model**

ปรเมษฐ์ รัตนานนท์^{1,3} ชัยกร ยิ่งเสรี¹ วรพล พงษ์เพ็ชร² และ ธนภัทร มั่งคะจิตร³

¹ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ปทุมธานี

²คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ กรุงเทพฯ

³สาขาวิศวกรรมข้อมูลขนาดใหญ่ คณะวิทยาลัยนวัตกรรมการด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัย
ธุรกิจบัณฑิต

ABSTRACT – The investment in the stock market has been interesting for domestic and foreign investors. There are many analysts attempting to improve forecasting models to increase prediction accuracy and stock return. Mostly focus on the single classifier using stock prices data as the main factor. However, the accuracy and stock return of the prediction were low. Thus, the purpose of this study attempt to increase prediction accuracy and stock return. Four factors were considered: stock prices data, indicator data, holding days and indicator days (the days that used to calculate indicators). In addition, 4 techniques of machine learning and ensemble model were used to forecast the trends of stock price by using SET's information from January 2011 to December 2016. The results show that ensemble model in term of weight form can increase the prediction accuracy by 5% - 14% and increase the stock return by 1% - 3%. Moreover, the factors of indicators data and holding days are important factors to improve efficiency of prediction and stock return.

KEY WORDS – stock; prediction; machine learning; ensemble model; SET

บทคัดย่อ – การลงทุนในตลาดหลักทรัพย์เป็นเรื่องที่น่าสนใจของเหล่านักลงทุนทั้งในและต่างประเทศ มีนักวิเคราะห์มากมายพยายามปรับปรุงโมเดลการพยากรณ์เพื่อเพิ่มประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุน แต่อย่างไรก็ตามพบว่างานวิจัยส่วนมากจะใช้ข้อมูลราคาซื้อขายหลักทรัพย์ (stock prices data) เป็นปัจจัยหลักในการสร้างโมเดลและใช้โมเดลเชิงเดียวในการวิเคราะห์ แต่ผลการพยากรณ์ดังกล่าวยังไม่มีความถูกต้องมากพอ ดังนั้นงานวิจัยนี้มีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพการพยากรณ์และผลตอบแทนการลงทุน โดยศึกษาปัจจัย 4 ปัจจัยคือ ข้อมูลราคาซื้อขายหลักทรัพย์, ข้อมูลดัชนีชี้วัดทางเทคนิค(indicators data), จำนวนวันที่ถือ(holding days) และจำนวนวันที่ใช้ในการคำนวณค่าดัชนีชี้วัดทางเทคนิค (indicator days) นอกจากนี้ได้ประยุกต์ใช้เทคนิคการวิเคราะห์การเรียนรู้ของเครื่อง (Machine Learning) 4 เทคนิค และเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Model) เพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ โดยใช้ข้อมูลของตลาดหลักทรัพย์แห่งประเทศไทยในช่วง มกราคม 2554 - ธันวาคม 2559 จากผลการวิจัยพบว่าเทคนิค Ensemble Model ด้วยวิธีการถ่วงน้ำหนักสามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ถึง 5% - 14 % เพิ่มผลตอบแทนการลงทุน 1% - 3% นอกจากนี้ข้อมูลดัชนีชี้วัดทางเทคนิคและจำนวนวันที่ถือเป็นปัจจัยสำคัญในการเพิ่มประสิทธิภาพในการพยากรณ์และผลตอบแทนในการลงทุน

คำสำคัญ – หุ้น; การพยากรณ์; การเรียนรู้ของเครื่อง; การเรียนรู้แบบรวมกลุ่ม; ตลาดหลักทรัพย์แห่งประเทศไทย

1. บทนำ

การลงทุนในหลักทรัพย์เป็นเรื่องที่น่าสนใจของนักลงทุนทั้งในและต่างประเทศ เพราะสามารถให้ผลตอบแทนที่สูงกว่าการฝากเงินไว้ในธนาคาร แต่เนื่องจากราคาหลักทรัพย์มีความผันผวนไม่แน่นอน ทำให้นักลงทุนมีความเสี่ยงที่อาจจะสูญเสียเงินลงทุน เมื่อราคาหลักทรัพย์ลดลงต่ำกว่าราคาที่นักลงทุนเข้าซื้อ ความเสี่ยงดังกล่าว ส่งผลให้นักลงทุนเกิดความจำเป็นที่จะต้องนำหลักการวิเคราะห์พยากรณ์ต่างๆ เข้ามาประยุกต์ใช้เพื่อพยากรณ์ทิศทางแนวโน้มการเปลี่ยนแปลงของราคาหลักทรัพย์ และกำหนดจังหวะในการเข้าซื้อขาย เพื่อลดความเสี่ยงของการขาดทุนที่อาจจะเกิดขึ้น

แม้ว่ามีงานวิจัยมากมายที่พยายามพัฒนาวิธีการพยากรณ์ราคาหลักทรัพย์โดยอาศัยหลักการจากสาขาวิชาต่างๆ ไม่ว่าจะเป็นฮิดเดนมาร์คอฟโมเดล[11], Random Forest[7], Neural Network[9], Support Vector Machine[12] และ Ensemble Model[2] อย่างไรก็ตามงานวิจัยส่วนมากจะมุ่งเน้นการศึกษาโมเดลเชิงเดี่ยวและใช้ข้อมูลราคาซื้อขายหลักทรัพย์รายวันในการสร้างโมเดล มีงานวิจัยไม่มากนักที่นำข้อมูลดัชนีชี้วัด (Indicator) ไปใช้ในการสร้างโมเดลการวิเคราะห์และนำการเรียนรู้แบบรวมกลุ่ม(Ensemble model) มาใช้เพื่อประสานการวิเคราะห์หลายรูปแบบเข้าด้วยกัน เพื่อให้ผลการวิเคราะห์มีความถูกต้องมากขึ้น

งานวิจัยฉบับนี้ทดสอบประสิทธิภาพการพยากรณ์และผลตอบแทนการลงทุนระหว่างโมเดลเชิงเดี่ยวกับการเรียนรู้แบบรวมกลุ่ม โดยใช้ข้อมูลหลักทรัพย์ 42 บริษัทที่อยู่ในดัชนี SET 50 ของตลาดหลักทรัพย์แห่งประเทศไทยบนข้อมูล 2 แบบคือ ข้อมูลราคาซื้อขายหลักทรัพย์ และข้อมูลดัชนีชี้วัดทางเทคนิคเพื่อเปรียบเทียบประสิทธิภาพ และผลตอบแทนการลงทุน โดยผลการทดสอบแสดงให้เห็นว่า โมเดลการเรียนรู้แบบกลุ่มนั้นมีความถูกต้องโดยเฉลี่ยสูงกว่าการพยากรณ์ของโมเดลเชิงเดี่ยว และค่าความถูกต้องโดยเฉลี่ยที่ได้จากการวิเคราะห์ข้อมูลดัชนีชี้วัดทางเทคนิคสูงกว่าการวิเคราะห์ข้อมูลราคาซื้อขายหลักทรัพย์รวมไปถึงผลตอบแทนการลงทุนด้วย

2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

การพยากรณ์(Prediction) เป็นสาขาหนึ่งของการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งแบ่งออกเป็น 2 รูปแบบหลักๆ คือ การพยากรณ์ด้วยการวิเคราะห์การถดถอย (Regression-base prediction) เป็นวิธีการหนึ่งที่ศึกษาความสัมพันธ์ระหว่างตัวแปร หรือคุณลักษณะของข้อมูล 2 ประเภท และนำผลของความสัมพันธ์ที่ได้จากการวิเคราะห์ไปพยากรณ์ ซึ่งผลของการวิเคราะห์จะอยู่ในรูปของข้อมูลเชิงปริมาณ (Quantitative data) เช่น อัตราผลตอบแทนของหลักทรัพย์หรือราคาหลักทรัพย์ในอนาคต ตัวอย่างงานวิจัยประเภทนี้ได้แก่ งานของ S.Nootyaskool *et al.*[11] ได้นำฮิดเดนมาร์คอฟโมเดลร่วมกับราคาหลักทรัพย์ในรูปแบบกราฟแท่งในการพยากรณ์การเปลี่ยนแปลงของหลักทรัพย์ ผลลัพธ์การทดลองให้ความถูกต้องในระดับหนึ่ง สะท้อนให้เห็นถึงความจำเป็นต้องใช้ปัจจัยอื่นๆ มาเรียนรู้เพิ่มในโมเดล

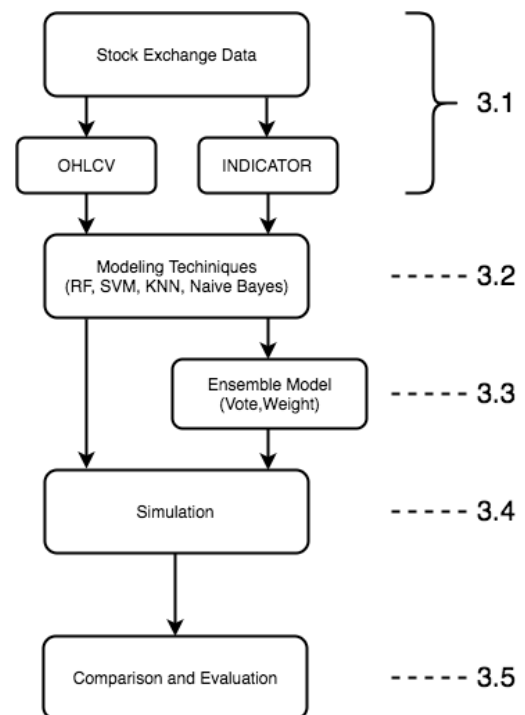
การพยากรณ์ด้วยการจำแนกข้อมูล (Classification-base prediction) เป็นกระบวนการสร้างโมเดลเรียนรู้ผลจากข้อมูลที่มีอยู่ โดยการสร้างกฎหรือเงื่อนไขเพื่อช่วยในการตัดสินใจจากข้อมูล เพื่อใช้ในการพยากรณ์ผลที่จะเกิดขึ้นจากข้อมูลในอนาคต ซึ่งผลของการวิเคราะห์ที่ได้จะอยู่ในรูปของข้อมูลเชิงคุณภาพ (Qualitative Data) เช่น การพยากรณ์ว่าราคาหลักทรัพย์จะปรับตัวสูงขึ้นหรือลดลง มีงานวิจัยหลายฉบับที่นำหลักการนี้ไปพยากรณ์ราคาหลักทรัพย์ เช่น งานวิจัยของ P.Sutheebanjard and W.Premchaiswadi[9] และ S.Chaigusin *et al.*[10] ได้นำโครงข่ายประสาท (Neural Network) ไปสร้างโมเดลพยากรณ์ราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย โดยมีกระบวนการฝึกฝนแบบ Supervise และใช้ขั้นตอนส่งค่าย้อนกลับ (backpropagation) ในการฝึกฝน ซึ่งผลลัพธ์ที่ได้มีค่าความผิดพลาดน้อยกว่า 2% งานวิจัยของ L.Khaidem *et al.*[7] ได้นำ Random Forest ไปใช้ในการพยากรณ์ราคาหลักทรัพย์เพื่อเปรียบเทียบการพยากรณ์กับวิธีอื่นๆ ซึ่งผลลัพธ์แสดงให้เห็นว่า Random Forest ให้ประสิทธิภาพดีที่สุด

นอกจากจะมีงานวิจัยจำนวนมากที่พยายามสร้างวิธีการต่างๆ เพื่อพยากรณ์ราคาหลักทรัพย์แล้ว ยังมีงานวิจัยบางกลุ่มพยายามนำเสนอวิธีการประยุกต์ใช้ Ensemble Model เพื่อพยากรณ์ราคาหลักทรัพย์ เช่น งานวิจัยของ B.Narayanan and M.Govindarajan[2] ได้นำเทคนิค Adaboost ซึ่งเป็นหนึ่งในเทคนิคการเรียนรู้แบบรวมกลุ่มมาเพิ่มประสิทธิภาพโมเดลพยากรณ์ราคาหลักทรัพย์ที่สร้างจาก Support Vector Machine และ Naïve Bayes ซึ่งผลการทดลองสรุปได้ว่า Ensemble Model ช่วยเพิ่มประสิทธิภาพความแม่นยำให้มากขึ้น งานวิจัยของ W.Qiang *et al.*[13] ได้นำ Ensemble Model มาช่วยลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของโมเดลพยากรณ์ราคาหลักทรัพย์ที่สร้างจาก Support Vector Machine และ Artificial Neural Networks ซึ่งผลการทดลองสรุปได้ Ensemble Model สามารถช่วยลดค่าความคลาดเคลื่อนกำลังสองได้

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่างานวิจัยจะมุ่งเน้นในการนำ Ensemble Model ไปช่วยเพิ่มประสิทธิภาพในแต่ละวิธีการจำแนกข้อมูล แต่ไม่ได้นำ Ensemble Model ไปใช้เพื่อรวมผลลัพธ์การพยากรณ์ และไม่มีการจำลองสถานการณ์

การลงทุนเพื่อให้เห็นผลตอบแทนการลงทุนตามผลการพยากรณ์ ดังนั้นงานวิจัยฉบับนี้จะทดสอบประสิทธิภาพการพยากรณ์และผลตอบแทนการลงทุนของโมเดลเชิงเดี่ยวกับ Ensemble Model กับข้อมูลหลักทรัพย์ 42 บริษัทที่อยู่ในดัชนี SET 50 ของตลาดหลักทรัพย์แห่งประเทศไทยบนข้อมูล 2 แบบ คือ ข้อมูลราคาซื้อขายหลักทรัพย์ และข้อมูลดัชนีชี้วัดทางเทคนิคเพื่อเปรียบเทียบประสิทธิภาพ และผลตอบแทนการลงทุน โดยผลการทดสอบแสดงให้เห็นว่า โมเดลการเรียนรู้แบบกลุ่มนั้นมีความถูกต้องโดยเฉลี่ยสูงกว่าการพยากรณ์ของโมเดลเชิงเดี่ยว และค่าความถูกต้องโดยเฉลี่ยที่ได้จากการวิเคราะห์ข้อมูลดัชนีชี้วัดทางเทคนิคสูงกว่าการวิเคราะห์ข้อมูลราคาซื้อขายหลักทรัพย์รวมไปถึงผลตอบแทนการลงทุนด้วย

3. วิธีการดำเนินวิจัย



รูปที่ 1. ขั้นตอนการวิเคราะห์ราคาหลักทรัพย์

3.1 การเตรียมข้อมูล

ชุดของข้อมูลหลักทรัพย์ที่ใช้ในการศึกษารวบรวมมาจากตลาดหลักทรัพย์แห่งประเทศไทย(The Stock Exchange of Thailand: SET) โดยกลุ่มข้อมูลแบ่งเป็น 2 กลุ่ม คือ 1. กลุ่มข้อมูลราคาซื้อขายหลักทรัพย์รายวัน ซึ่งประกอบไปด้วย ราคาเปิด ราคาสูงสุด ราคาต่ำสุด ราคาปิด และ ปริมาณการซื้อขาย(open, high, low, close, volume) 2. กลุ่มข้อมูลดัชนีชี้วัดทางเทคนิค(indicator technical) ตามตารางที่ 1 และ 2

ตารางที่ 1 แสดงตัวอย่างข้อมูลราคาซื้อขายหลักทรัพย์

TICKER	OPEN	HIGH	LOW	CLOSE	VOL
stock_01	3.02	3.08	3.02	3.06	18455500
stock_01	3.06	3.10	3.04	3.06	45132800
stock_01	3.04	3.06	2.96	3.00	52636700
stock_01	3.02	3.08	3.02	3.02	41214700

ตารางที่ 2 แสดงตัวอย่างข้อมูลดัชนีชี้วัดทางเทคนิค

TICKER	STO_K	STO_D	OBV	RSI	WPR	Volatility	ROC
stock_01	0.73	0.73	12945246566	69.55	0.26	0.33	0.08
stock_01	0.14	0.57	12832396366	49.44	0.85	0.39	-0.01
stock_01	0.12	0.33	12796090966	44.13	0.87	0.37	-0.03
stock_01	0.37	0.21	12899707966	55.96	0.57	0.31	-0.02
stock_01	0.60	0.46	12899707966	55.96	0.40	0.19	-0.01

Stochastic Oscillator (STO) เป็นการวิเคราะห์ที่ช่วยให้รู้ว่า ราคาปิดของหลักทรัพย์อยู่ในสภาพใดเมื่อเปรียบเทียบกับราคาหลักทรัพย์ในช่วงเวลาที่เราส่ง STO จะเคลื่อนไหวตามแรงส่ง (โมเมนตัม: Momentum) ของราคา ไม่ได้เคลื่อนไหวไปตามแนวโน้ม หรือปริมาณการซื้อขาย ซึ่งเหมาะกับการวิเคราะห์ตลาด Sideway (ตลาดที่มีลักษณะที่ราคาหลักทรัพย์ไม่ขึ้นไม่ลง หรือขึ้นลงในกรอบราคาแคบๆ) โดยจะมีการแสดงผลเป็น 2 เส้น ได้แก่ เส้นหลักเรียกว่า %K และ เส้นค่าเฉลี่ยเคลื่อนที่ของ %K เรียกว่า %D ซึ่งสามารถหาได้จากสมการ (1) และ (2)

$$\%K = \frac{C-L(n)}{H(n)-L(n)} \times 100 \quad (1)$$

$$\%D = 3 - \text{period moving average of \%K} \quad (2)$$

เมื่อ C = ราคาปิด ณ วันปัจจุบัน

H(n) = ราคาสูงสุดในช่วงเวลา n วันที่ผ่านมา

L(n) = ราคาต่ำสุดในช่วง n วันที่ผ่านมา

Relative Strength Index (RSI) เป็นดัชนีชี้วัดที่ใช้วัดการเปลี่ยนแปลงการเคลื่อนไหวของราคาหลักทรัพย์ที่เกิดขึ้นเพื่อดูภาวะการซื้อมากเกินไป (Overbought) หรือขายมากเกินไป (Oversold) ซึ่ง RSI จะมีค่าอยู่ระหว่าง 0-100 โดยที่ RSI มากกว่า 70 จะแสดงถึงสภาวะ Overbought มีโอกาสที่ราคาจะปรับตัวลงมา และ RSI น้อยกว่า 30 จะแสดงถึงสภาวะ Oversold ราคามีโอกาสที่จะปรับตัวสูงขึ้นได้ สามารถคำนวณได้จากสมการ (3) และ (4)

$$RSI = 100 - \frac{100}{1+RS} \quad (3)$$

$$RS = \frac{\text{Average Gain Over past (n) day}}{\text{Average Loss Over past (n) day}} \quad (4)$$

เมื่อ Average Gain Over past (n) day = ราคาปิดบวกเฉลี่ยในช่วงเวลา n วันที่ผ่านมา

Average Loss Over past (n) day = ราคาปิดลบเฉลี่ยในช่วง n วันที่ผ่านมา

On Balance Volume (OBV) เป็นดัชนีชี้วัดปริมาณการซื้อขายและการเปลี่ยนราคา มีหลักการคิดเบื้องต้นคือ ปริมาณของหลักทรัพย์วันที่ปิดบวก (ราคาปิดสูงกว่าราคาเปิด) จะถือว่าเป็นการสะสมหลักทรัพย์ ส่วนปริมาณของหลักทรัพย์วันที่ปิดลบ (ราคาปิดต่ำกว่าราคาเปิด) จะเป็นการกระจายหลักทรัพย์ สามารถคำนวณได้จากสมการ (5)

$$OBV(n) = \begin{cases} OBV(n-1) + Vol(n) & \text{if } C(n) > C(n-1) \\ OBV(n-1) - Vol(n) & \text{if } C(n) < C(n-1) \\ OBV(n-1) & \text{if } C(n) = C(n-1) \end{cases} \quad (5)$$

เมื่อ OBV(n) = On Balance Volume ช่วงเวลา n วันที่ผ่านมา

Vol(n) = ปริมาณซื้อขาย ในช่วงเวลา n วันที่ผ่านมา

C(n) = ราคาปิดในช่วงเวลา n วันที่ผ่านมา

Williams Percent Range (WPR) เป็นดัชนีชี้วัดที่มีการเปลี่ยนแปลงเสมอ (dynamic) ซึ่งใช้ในการกำหนดสภาวะของหลักทรัพย์ว่าอยู่ในสภาวะ Overbought หรือ Oversold ดัชนีชี้วัด WPR มีลักษณะการใช้คล้ายกับ STO จะต่างกันแค่ WPR จะมีสเกลแบบกลับทิศทางกัน โดยที่ถ้าค่าดัชนีชี้วัดอยู่ระหว่าง -80% ถึง -100% หลักทรัพย์จะอยู่ในสภาวะ Oversold และถ้าค่าดัชนีชี้วัดอยู่ระหว่าง 0% ถึง -20% จะอยู่ในสภาวะ Overbought สามารถคำนวณหาค่าดัชนีชี้วัดได้จากสมการ (6)

$$\%R = \frac{(H(n)-C)}{(H(n)-L(n))} \times -100 \quad (6)$$

เมื่อ C = ราคาปิด ณ วันปัจจุบัน

H(n) = ราคาสูงสุดในช่วงเวลา n วันที่ผ่านมา

L(n) = ราคาต่ำสุดในช่วง n วันที่ผ่านมา

ความผันผวน (VOLATILITY) เป็นปัจจัยหนึ่งที่มีความสำคัญของนักลงทุนในการเข้าลงทุนในหลักทรัพย์นั้นๆ หลักทรัพย์ที่มี Volatility สูงราคาหลักทรัพย์นั้นก็จะมีการผันผวนมาก ซึ่งหมายความว่าหลักทรัพย์ตัวนั้นมีความเสี่ยงสูง ในการวิเคราะห์หา Volatility มีได้หลายวิธีการเช่น Close-to-Close Volatility, OHLC Volatility: Garman and Klass และ High-Low Volatility: Parkinson เป็นต้น ซึ่งในงานวิจัยฉบับนี้ได้ใช้วิธี

Close-to-Close Volatility สามารถคำนวณได้จากสมการ (7) (8)

และ (9)

$$\sigma_{cl} = \sqrt{\frac{N}{n-2} \sum_{i=1}^{n-1} (r(i) - \bar{r})^2} \quad (7)$$

$$r(i) = \log \left(\frac{C(i)}{C(i-1)} \right) \quad (8)$$

$$\bar{r} = \frac{r_1 + r_2 + \dots + r_{n-1}}{n-1} \quad (9)$$

เมื่อ

σ = ค่าความผันผวน

n = จำนวนวันในการประมาณความผันผวน

N = จำนวนวันต่อปี

C(i) = ราคาปิดในวันที่ i

r(i) = ผลตอบแทนในวันที่ i

Rate of Change (ROC) หรืออัตราค่าการเปลี่ยนแปลงของราคาจะแสดงถึงความแตกต่างระหว่างราคาปัจจุบันกับราคาย้อนหลัง ซึ่งช่วยในการกำหนดสถานะของหลักทรัพย์ ซึ่งค่า ROC สูงยิ่งแสดงถึงเข้าสภาวะ Overbought และค่า ROC ต่ำจะแสดงถึงการเข้าสภาวะ Oversold เท่านั้น สามารถคำนวณหา ROC ได้จากสมการ (10)

$$ROC = \frac{C - C(n)}{C(n)} \times 100 \quad (10)$$

เมื่อ

C = ราคาปิด ณ วันปัจจุบัน

C(n) = ราคาปิดในช่วงเวลา n วันที่ผ่านมา

โดยจำนวนวันที่ใช้ในการคำนวณค่าดัชนีชี้วัด (n) ที่ใช้ในการวิจัยฉบับนี้มี 20 ค่า แต่ละค่ามีช่วงห่างจำนวน 5 วัน ได้แก่ 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 และ 100 วัน

ข้อมูลหลักทรัพย์ตราสารทุนในปัจจุบันมีประมาณ 2,011 ตัว ในการวิจัยนี้ใช้กลุ่มข้อมูลหลักทรัพย์จำนวน 42 หลักทรัพย์ที่อยู่ในดัชนี SET 50 ตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2559 ซึ่งข้อมูลจะถูกวิเคราะห์ 4 ปีได้แก่ปี 2556, 2557, 2558 และ 2559 ซึ่งข้อมูลที่ใช้ในการวิเคราะห์ในแต่ละปีแบ่งออกเป็น 3 ส่วน ได้แก่ กลุ่มข้อมูลทดลอง(train data) จำนวน 24 เดือน กลุ่มข้อมูลทดสอบ(test data) จำนวน 6 เดือน และกลุ่มข้อมูลประเมินผล(evaluation data) จำนวน 6 เดือน ตามตารางที่ 3

ตาราง 3 แสดงการแบ่งข้อมูลเพื่อการวิเคราะห์

ปีที่วิเคราะห์	กลุ่มข้อมูล		
	ข้อมูลทดลอง (24 เดือน)	ข้อมูลทดสอบ (6 เดือน)	ข้อมูลวิเคราะห์ (6 เดือน)
2556	1 ม.ค. 2554 ถึง 31 ธ.ค. 2555	1 ม.ค. 2556 ถึง 30 มิ.ย. 2556	1 ก.ค. 2556 ถึง 31 ธ.ค. 2556
2557	1 ม.ค. 2555 ถึง 31 ธ.ค. 2556	1 ม.ค. 2557 ถึง 30 มิ.ย. 2557	1 ก.ค. 2557 ถึง 31 ธ.ค. 2557
2558	1 ม.ค. 2556 ถึง 31 ธ.ค. 2557	1 ม.ค. 2558 ถึง 30 มิ.ย. 2558	1 ก.ค. 2558 ถึง 31 ธ.ค. 2558
2559	1 ม.ค. 25557 ถึง 31 ธ.ค. 2558	1 ม.ค. 2559 ถึง 30 มิ.ย. 2559	1 ก.ค. 2559 ถึง 31 ธ.ค. 2559

3.2 Modeling Techniques

เป็นกระบวนการสร้างโมเดลเรียนรู้จากข้อมูลที่มีอยู่ โดยการสร้างกฎหรือเงื่อนไขเพื่อช่วยในการตัดสินใจจากข้อมูล เพื่อใช้ในการพยากรณ์ผลที่จะเกิดขึ้นจากข้อมูลในอนาคต ซึ่งในงานวิจัยนี้ได้ใช้ 4 เทคนิคในการสร้างโมเดลคือ Random Forest, Support Vector Machine, K nearest neighbors และ Naïve Bayes โดยใช้ข้อมูล 2 กลุ่มด้วยกันคือ ข้อมูลราคาซื้อขายหลักทรัพย์ และข้อมูลดัชนีชี้วัดทางเทคนิค ในการสร้างโมเดลพยากรณ์เพื่อวัดค่าประสิทธิภาพจากเทคนิคการจำแนกประเภทข้อมูล 4 เทคนิค รายละเอียดของเทคนิคที่ใช้เป็นดังนี้

3.2.1 Random Forest

แนวคิดของ Random Forest[8]นี้คือการสร้างโมเดลด้วยวิธีการ Decision Tree ขึ้นมาหลายๆ โมเดลโดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละโมเดลมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย วิธีการของ Decision Tree คือเทคนิคที่ให้ผลลัพธ์ในลักษณะเป็นโครงสร้างของต้นไม้ ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัวทดสอบ กิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (class) ก็คือผลลัพธ์ที่ได้จากการพยากรณ์ ซึ่งข้อดีของวิธีการนี้คือให้ผลการพยากรณ์ที่แม่นยำและเกิดปัญหา overfitting น้อย

3.2.2 Support Vector Machine

แนวคิดหลักของ Support Vector Machine[8] คือการหาระนาบการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยจะพยายามสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงระหว่างกึ่งกลางของกลุ่มข้อมูลให้มีระยะห่างระหว่างขอบเขตของทั้งสองกลุ่มมากที่สุด โดยปกติ Support Vector Machine ถูกนำมาใช้กับข้อมูลเชิงเส้น แต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้มักเป็นข้อมูลแบบไม่เชิงเส้น ซึ่งสามารถแก้ไขปัญหาดังกล่าวด้วยการนำ Kernel Function มาใช้ ซึ่งใน Support Vector Machine จะมี Kernel Function ให้ผู้ใช้สามารถประยุกต์ใช้ในการแก้ปัญหาก็ได้หลายวิธี เช่น Linear, Polynomial, Radial Basis Function โดยงานวิจัยนี้ได้ใช้ Radial Basis Function ในการทดลอง ข้อดีของวิธีการนี้คือเกิดปัญหา overfitting น้อยและมี Kernel Function ในการวิเคราะห์ข้อมูลแบบไม่เชิงเส้นได้

3.2.3 K nearest neighbors

K nearest neighbors[3,4]เป็นวิธีการหนึ่งสำหรับการจำแนกกลุ่มของข้อมูลที่มีลักษณะกระจัดกระจาย โดยจะหาระยะห่างของข้อมูลคุณลักษณะที่ต้องการพยากรณ์กับข้อมูลผลลัพธ์ที่มีอยู่ก่อนหน้า โดยจะเลือกข้อมูลที่ระยะห่างที่ใกล้ที่สุดจำนวน K ตัว จากนั้นทำการสกัดข้อมูล จำแนกและพยากรณ์ผลจากการข้อมูลที่ได้มา ข้อดีของวิธีการนี้คือสามารถให้ผลประสิทธิภาพที่ดีถ้า training data มีขนาดใหญ่

3.2.4 Naïve Bayes

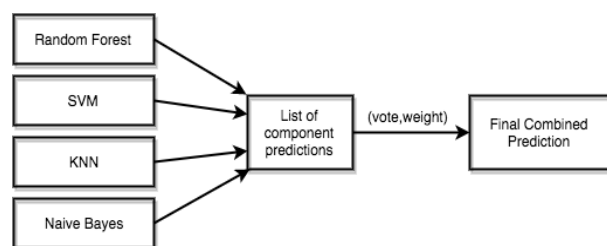
Naïve Bayes[8] เป็นการจำแนกประเภทข้อมูลที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของทฤษฎีของเบย์ (Bayes' Theorem) และสมมติฐานที่ให้การเกิดของเหตุการณ์ต่างๆเป็นอิสระต่อกัน (Independence) ถ้ากำหนดให้ $P(h)$ คือความน่าจะเป็นที่จะเกิดเหตุการณ์ h และ $P(h|D)$ คือความน่าจะเป็นที่จะเกิดเหตุการณ์ h เมื่อเกิดเหตุการณ์ D จากตัวแปรที่กำหนดและแนวคิดของ Bayes' Theorem นั้นสามารถทำนายเหตุการณ์ที่พิจารณาจากการเกิดของเหตุการณ์ต่างๆ ข้อดีของวิธีการนี้คือเข้าใจง่ายและง่ายต่อการนำไปใช้งาน

ในงานวิจัยฉบับนี้ในแต่ละเทคนิคได้สร้างโมเดล

แยกตามจำนวนหลักทรัพย์ จำนวนวันที่ถือ(Holding day) และจำนวนวันดัชนีย้อนหลัง(indicator day) เป็นจำนวน 50,400 โมเดล แล้วนำผลการทดลองที่ได้ไปหาค่าเฉลี่ยความถูกต้องและนำไปจำลองการลงทุนในตลาดหลักทรัพย์เพื่อคำนวณค่าเฉลี่ยผลตอบแทนการลงทุนของแต่ละเทคนิค

3.3 Ensemble Model

Ensemble Model[14] เป็นวิธีการที่อาศัยโมเดลการจำแนกข้อมูลมากกว่าหนึ่งตัว (Base Classification) ซึ่งแต่ละโมเดลการจำแนกข้อมูลจะมีกระบวนการทำงานของตัวเอง และทุกตัวของโมเดลการจำแนกจะสร้างจากกลุ่มข้อมูลเดียวกัน เมื่อได้ผลลัพธ์ของแต่ละโมเดล ก็จะนำผลลัพธ์ที่ได้เหล่านั้นผ่านวิธีการรวบรวม (Combination, Integration หรือ Vote) และตัดสินใจผลลัพธ์สุดท้าย (Final Decision) เพื่อให้ได้ผลลัพธ์การจำแนกข้อมูลเดียวกัน



รูปที่ 2. ขั้นตอนการรวมโมเดลพยากรณ์

การรวมโมเดลพยากรณ์จะใช้วิธีการบูสต์ติง (Boosting) มีหลักการคือจะทำการสร้างโมเดลจำแนกประเภทข้อมูลหลายโมเดล แต่ละโมเดลจะใช้ train data ชุดเดียวกันในการสร้าง ซึ่งแต่ละโมเดลจะมีค่าถ่วงน้ำหนัก (Weight) เพิ่มเข้ามา โดยค่าถ่วงน้ำหนักนี้ได้มาจากความถูกต้องของการเรียนรู้บนชุดข้อมูลสำหรับคำตอบสุดท้ายของการทำงานด้วยวิธีบูสต์ติงนั้น จะใช้วิธีการโหวตแบบถ่วงน้ำหนักแล้วกำหนดกลุ่มให้ข้อมูลใหม่ด้วยผลโหวตที่มากที่สุด (Majority Voting) ซึ่งในงานวิจัยฉบับนี้ได้ใช้วิธีรวม 2 แบบ คือ แบบค่าเฉลี่ย และ แบบถ่วงน้ำหนัก

ค่าถ่วงน้ำหนัก (weight) เป็นค่าที่ได้จากการแปลงค่าความถูกต้องจากการทดสอบพยากรณ์ของ test data ในแต่ละโมเดล ดังในสมการที่ (11)

$$Weight(i) = \frac{Accuracy(i)}{100} \quad (11)$$

โดยที่ Weight(i) คือ ค่าถ่วงน้ำหนักของโมเดล i
Accuracy(i) คือ ค่าความถูกต้องของโมเดล i จากข้อมูล test data (%)

3.4 การจำลองการลงทุนในตลาดหลักทรัพย์

ขั้นตอนการจำลองการลงทุน เพื่อนำผลที่ได้จากการจำแนกข้อมูล 4 เทคนิค และ Ensemble Model มาทดลองลงทุนเพื่อหาผลตอบแทนการลงทุนตามจำนวนวันที่ถือ(Holding day) ได้จำลองการถือหลักทรัพย์ตั้งแต่ถือ 1 วันถึงถือ 60 วัน แล้วนำผลตอบแทนการลงทุนที่ได้ในแต่ละวันที่ถือไปคำนวณหาค่าเฉลี่ยผลตอบแทนการลงทุนโดยรวม โดยเริ่มซื้อหลักทรัพย์เมื่อตลาดเปิด และขายหลักทรัพย์ก่อนตลาดปิดในวันที่กำหนดผลตอบแทนการลงทุนตามวันที่ถือสามารถคำนวณได้จากสมการที่ (12) โดยในงานวิจัยนี้ไม่ได้คำนวณอัตราค่าธรรมเนียมการซื้อขายหลักทรัพย์

$$Profit = \frac{O-C(n)}{O} \times 100 \quad (12)$$

เมื่อ O คือ ราคาเปิด ณ วันปัจจุบัน

C(n) คือ ราคาปิด ณ วันที่ n

n คือ จำนวนวันที่ถือ

3.5 การวัดประสิทธิภาพ

ขั้นตอนการเปรียบเทียบโมเดลพิจารณาจากค่าความถูกต้อง (Accuracy) ดังในสมการที่ (13)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

โดยที่ True Positive(TP) คือ จำนวนที่โมเดลพยากรณ์ว่า True และในข้อมูลเป็น True

False Positive(FP) คือ จำนวนที่โมเดลพยากรณ์ว่า True แต่ในข้อมูลเป็น False

True Negative(TN) คือ จำนวนที่โมเดลพยากรณ์ว่า False และในข้อมูลเป็น False

False Negative(FN) คือ จำนวนที่โมเดลพยากรณ์ว่า False แต่ในข้อมูลเป็น True

โดยในงานวิจัยฉบับนี้ผลการทดลองที่ได้จะเป็นค่าเฉลี่ยความถูกต้องจากโมเดลสร้างจากข้อมูลราคาซื้อขายหลักทรัพย์จำนวน 2,520 โมเดล ต่อเทคนิคต่อปี และ โมเดลที่สร้างข้อมูลดัชนีชี้วัดทางเทคนิค 50,400 โมเดล ต่อเทคนิคต่อปี

4. ผลการทดลอง

4.1 ผลและวิเคราะห์การเปรียบเทียบโมเดลพยากรณ์แนวโน้มของราคาหลักทรัพย์

จากตาราง 4 การเปรียบเทียบประสิทธิภาพโมเดลการพยากรณ์แนวโน้มของราคาหลักทรัพย์ด้วยข้อมูลราคาซื้อขายจากเทคนิคการจำแนกประเภทข้อมูล 4 เทคนิคและ Ensemble Model นั้นพบว่า Ensemble Model แบบถ่วงน้ำหนักมีประสิทธิภาพในการพยากรณ์ได้ดีที่สุด โดยมีค่าความถูกต้องอยู่ที่ 51.09%, 57.73%, 55.72% และ 58.93% สำหรับการวิเคราะห์ในปี 2556, 2557, 2558 และ 2559 มีค่าความถูกต้องโดยเฉลี่ยรวมอยู่ที่ 55.87%

ตาราง 4. แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องของโมเดลพยากรณ์แนวโน้มราคาหลักทรัพย์ด้วยข้อมูลราคาซื้อขาย(%)

Model		Year				Average
		2556	2557	2558	2559	
RF		48.28%	53.49%	48.81%	53.47%	51.01%(2)
SVM		47.07%	51.56%	50.07%	54.47%	50.79%(4)
KNN		46.33%	51.76%	50.19%	51.52%	49.95%(6)
NB		49.27%	48.72%	49.73%	52.44%	50.04%(5)
EB	ew	46.93%	51.56%	48.98%	53.29%	50.95%(3)
	w	51.09%	57.73%	55.72%	58.93%	55.87%(1)

RF = Random Forest, SVM = Support Vector Machine, KNN = K-nearest Neighbor, NB = Naïve Bayes, B&H = Buy and Hold, EB = Ensemble, ew = equal weight, w = weight

จากตาราง 5 การเปรียบเทียบประสิทธิภาพโมเดลการพยากรณ์แนวโน้มของราคาหลักทรัพย์ด้วยข้อมูลดัชนีชี้วัดทางเทคนิคพบว่า Ensemble Model แบบถ่วงน้ำหนักมีประสิทธิภาพการพยากรณ์ได้ดีที่สุดโดยมีความถูกต้องโดยเฉลี่ยรวมอยู่ที่ 66.74% และในปี 2556, 2557 และ 2558 มีค่าความถูกต้องอยู่ที่ 69.53%, 68.05%, และ 67.41% ตามลำดับ แต่ในปี 2559 โมเดล Random Forest มีประสิทธิภาพในการพยากรณ์ได้ดีที่สุด มีค่าความถูกต้อง 62.23%

ตาราง 5. แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องของโมเดลพยากรณ์แนวโน้มราคาหลักทรัพย์ด้วยข้อมูลดัชนีชี้วัดทางเทคนิค(%)

Model		Year				Average
		2556	2557	2558	2559	
RF		68.90%	64.50%	65.10%	62.23%	65.18%(2)
SVM		61.91%	62.41%	64.06%	54.81%	60.80%(4)
KNN		53.44%	56.24%	49.78%	49.00%	52.12%(6)
NB		61.19%	54.69%	55.87%	53.69%	56.36%(5)
EB	ew	64.53%	62.58%	61.30%	57.06%	61.36%(3)
	w	69.53%	68.05%	67.41%	61.91%	66.74%(1)

4.2 ผลและวิเคราะห์การเปรียบเทียบผลตอบแทนจากการจำลองการลงทุน

จากตารางที่ 6 การจำลองสถานการณ์การลงทุนในตลาดหลักทรัพย์ตามโมเดลที่สร้างจากข้อมูลราคาหลักทรัพย์ พบว่า Ensemble Model แบบถ่วงน้ำหนักให้ผลตอบแทนการลงทุนโดยเฉลี่ยรวมได้ดีที่สุดอยู่ที่ 1.12% และในปี 2556 และ 2558 มีผลตอบแทนการลงทุนโดยเฉลี่ยอยู่ที่ 0.62% และ 0.48% ตามลำดับ แต่ในปี 2557 โมเดล NaiveBayes ให้ผลตอบแทนการลงทุนดีที่สุด มีผลตอบแทนการลงทุนอยู่ที่ 1.95% ในปี 2559 โมเดล Random Forest ให้ผลตอบแทนที่ดีที่สุด มีผลตอบแทนการลงทุนอยู่ที่ 3.25% เมื่อเปรียบเทียบกับ การซื้อแล้วถือระยะยาว(Buy-and-Hold: B&H) จะพบว่าผลตอบแทนการลงทุนโดยเฉลี่ยในทุกๆเทคนิคมีผลที่ดีกว่าผลของการซื้อและถือระยะยาว

ตาราง 6. แสดงผลการเปรียบเทียบผลตอบแทนการลงทุนด้วยโมเดลที่สร้างจากข้อมูลราคาซื้อขายหลักทรัพย์(%)

Model		Year				Average
		2556	2557	2558	2559	
RF		-1.19%	1.14%	-0.65%	3.25%	0.64%(4)
SVM		-1.60%	1.41%	-0.72%	1.62%	0.98%(2)
KNN		-1.74%	-0.87%	-1.61%	0.94%	-0.82%(6)
NB		-0.29%	1.95%	-0.29%	1.80%	0.79%(3)
B&H		-2.91%	-1.59%	-3.02%	1.92%	-1.40%(7)
EB	ew	-1.43%	1.38%	-0.79%	1.34%	0.13%(5)
	w	0.62%	1.42%	0.48%	1.96%	1.12%(1)

จากตารางที่ 7 การจำลองสถานการณ์การลงทุนในตลาดหลักทรัพย์ตามโมเดลที่สร้างจากข้อมูลดัชนีชี้วัดทางเทคนิคพบว่าในปี 2557 และ 2558 Ensemble Model แบบถ่วงน้ำหนักให้ผลตอบแทนการลงทุนโดยเฉลี่ยสูงสุด ได้แก่ 3.17% และ 3.02% ตามลำดับ ในปี 2556 และ 2559 โมเดล Random Forest ให้ผลตอบแทนการลงทุนดีที่สุด มีผลตอบแทนการลงทุนอยู่ที่ 4.56% และ 4.12% ตามลำดับ และผลตอบแทนการลงทุนโดยเฉลี่ย Random Forest ให้ผลดีที่สุดอยู่ที่ 3.64% เมื่อเปรียบเทียบกับ B&H จะพบว่าผลตอบแทนการลงทุนโดยเฉลี่ยในทุกๆเทคนิคมีผลที่ดีกว่าผลของ B&H

ตาราง 7. แสดงผลการเปรียบเทียบผลตอบแทนการลงทุนด้วยโมเดลที่สร้างจากข้อมูลดัชนีชี้วัดทางเทคนิค(%)

Model		Year				Average
		2556	2557	2558	2559	
RF		4.56%	3.29%	2.58%	4.12%	3.64%(1)
SVM		3.25%	1.03%	2.14%	1.74%	2.04%(4)
KNN		1.24%	1.23%	-1.01%	1.72%	0.80%(6)
NB		2.77%	1.40%	1.72%	1.90%	1.95%(5)
B&H		-2.91%	-1.59%	-3.02%	1.92%	-1.40%(7)
EB	ew	2.85%	2.64%	2.13%	3.00%	2.68%(3)
	w	3.74%	3.17%	3.02%	3.58%	3.38%(2)

ตาราง 8 แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุน จากโมเดลที่สร้างด้วยการเรียนรู้แบบรวมกลุ่มตามจำนวนวันที่ถือ

Holding day	Average Accuracy(%)				Average Profit(%)			
	OHLCV		INDICATOR		OHLCV		INDICATOR	
	EW	W	EW	W	EW	W	EW	W
1	53.72	54.72	54.64	56.41	-0.04	0.00	0.22	0.26
5	52.03	54.46	57.53	60.08	0.20	0.32	1.30	1.56
10	51.55	55.50	58.39	61.94	0.45	0.85	2.08	2.71
15	49.67	54.85	59.74	64.05	0.28	0.54	2.64	3.13
20	50.30	56.01	60.71	65.76	0.21	0.47	2.70	3.45
25	50.29	55.34	61.60	66.60	0.21	0.74	2.53	3.53
30	49.40	55.58	62.69	67.89	-0.01	0.63	2.75	3.78
35	49.55	55.82	62.56	68.59	-0.04	0.40	2.83	3.82
40	50.89	56.30	62.32	69.26	0.32	0.92	2.88	3.83
45	49.38	56.82	62.44	70.08	-0.07	0.46	3.06	3.84
50	50.79	57.96	62.94	69.97	0.69	1.31	3.49	4.23
55	50.14	57.51	63.04	69.28	0.15	0.52	4.02	4.70
60	48.73	59.04	61.62	69.27	0.24	0.94	3.94	5.01

4.3 ผลและวิเคราะห์การเปรียบเทียบประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุน จากโมเดลที่สร้างด้วยการเรียนรู้แบบรวมกลุ่ม ตามจำนวนวันที่ถือ

จากตารางที่ 8 พบว่าประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุนของ Ensemble Model ด้วยวิธีการถ่วงน้ำหนักมีแนวโน้มเพิ่มขึ้นตามจำนวนวันที่ถือ จากโมเดลที่สร้างด้วยข้อมูลราคาซื้อขายหลักทรัพย์ จำนวนวันที่ถือ 60 วัน ให้ค่าความถูกต้องดีที่สุดอยู่ที่ 59.04% และจำนวนวัน 50 วันให้ผลตอบแทนการลงทุนดีที่สุดอยู่ที่ 1.31%

จากโมเดลที่สร้างด้วยข้อมูลดัชนีชี้วัดทางเทคนิค จำนวนวันที่ถือ 45 วัน ให้ค่าความถูกต้องดีที่สุดอยู่ที่ 70.08% และ จำนวนวัน 60 วัน ให้ผลตอบแทนการลงทุนดีที่สุดอยู่ที่ 5.01 %

4.4 ผลและวิเคราะห์การเปรียบเทียบประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุน จากโมเดลที่สร้างด้วยการเรียนรู้แบบรวมกลุ่ม ตามจำนวนวันที่ใช้คำนวณดัชนีชี้วัดทางเทคนิค

จากตารางที่ 9 พบว่าผลตอบแทนการลงทุนของ Ensemble Model ด้วยวิธีการถ่วงน้ำหนักมีแนวโน้มเพิ่มขึ้นตามจำนวนวันที่ใช้คำนวณดัชนีชี้วัดทางเทคนิค โดยจำนวนวัน 10 วันให้

ค่าความถูกต้องดีที่สุดอยู่ที่ 67.51% และจำนวนวัน 100 วันให้ผลตอบแทนการลงทุนดีที่สุดอยู่ที่ 3.68%

ตาราง 9 แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องและผลตอบแทนการลงทุน จากโมเดลที่สร้างด้วยการเรียนรู้แบบรวมกลุ่ม ตามจำนวนวันที่ใช้คำนวณดัชนีชี้วัดทางเทคนิค

Indicator day	INDICATOR			
	Avg. Accuracy(%)		Avg. Profit(%)	
	EW	W	EW	W
10	61.97	67.51	2.74	3.37
20	60.59	66.52	2.45	3.19
30	60.61	65.87	2.47	3.16
40	60.68	66.26	2.27	3.06
50	60.76	66.22	2.42	3.21
60	60.92	66.24	2.61	3.28
70	61.21	66.25	2.71	3.40
80	61.47	66.54	2.71	3.49
90	61.27	66.41	2.85	3.50
100	62.58	67.05	2.98	3.68

5. สรุปผลการทดลอง

งานวิจัยนี้ได้ศึกษาความถูกต้องและจำลองการลงทุนเพื่อหาผลตอบแทนจากการพยากรณ์แนวโน้มของราคาหลักทรัพย์รายวันในดัชนี SET 50 โดยกลุ่มข้อมูลที่ศึกษาจะถูกแบ่งเป็น 2 กลุ่มคือ กลุ่มข้อมูลราคาซื้อขายหลักทรัพย์รายวัน และกลุ่ม

ข้อมูลดัชนีชี้วัดทางเทคนิค เทคนิคการจำแนกประเภทข้อมูล 4 เทคนิคและ Ensemble Model เพื่อรวมรูปแบบการวิเคราะห์ทุกโมเดลเข้าด้วยกัน ซึ่งทำการวิเคราะห์ข้อมูลของตลาดหลักทรัพย์แห่งประเทศไทย ในช่วงเดือนมกราคม พ.ศ. 2556 ถึงเดือนธันวาคม พ.ศ. 2559

จากผลการทดลอง แสดงให้เห็นว่าข้อมูลดัชนีชี้วัดทางเทคนิคและจำนวนวันที่ถือเป็นปัจจัยสำคัญที่ช่วยในการเพิ่มประสิทธิภาพโมเดลการพยากรณ์แนวโน้มของราคาหลักทรัพย์ ข้อมูลดัชนีชี้วัดทางเทคนิคสามารถเพิ่มค่าความถูกต้องได้ถึง 2% - 14% รวมไปถึงเพิ่มผลตอบแทนการลงทุนได้ถึง 1% - 3% และจำนวนวันที่ถือสามารถเพิ่มค่าความถูกต้องได้ถึง 5%-15% รวมไปถึงเพิ่มผลตอบแทนการลงทุนได้ถึง 1% - 5%

จำนวนวันที่ใช้ในการคำนวณดัชนีชี้วัดทางเทคนิคมีผลต่อผลตอบแทนการลงทุนโดยสามารถเพิ่มได้ 0.1% - 0.5 %

จากการนำเสนอเทคนิคการจำแนกประเภทข้อมูล 4 เทคนิคและ Ensemble Model เพื่อเพิ่มประสิทธิภาพในการพยากรณ์แนวโน้มของราคาหลักทรัพย์ ผลการทดลองสามารถสรุปได้ว่า Ensemble Model ด้วยวิธีการถ่วงน้ำหนักสามารถเพิ่มประสิทธิภาพการพยากรณ์ให้ดีขึ้นได้ สามารถเพิ่มค่าความถูกต้องได้ถึง 1% - 14% ผลการทดลองของผลตอบแทนการลงทุนถึงแม้ว่า Random Forest จะมีผลตอบแทนการลงทุนมากที่สุดที่ 3.64% Ensemble Model ก็ให้ผลตอบแทนการลงทุนที่ 3.38% ซึ่งมีความแตกต่างไม่มาก แต่ Ensemble Model สามารถเพิ่มผลตอบแทนการลงทุนได้ถึง 1% - 3%

เอกสารอ้างอิง

- [1] A.C. Charu, Outlier Analysis, 2nd ed. New York: IBM T. J. Watson Research Center Yorktown Heights, 2016.
- [2] B. Narayanon and M. Govindarajan, "Prediction on Stock Market using Ensemble Model", International Journal of Computer Applications, vol. 128, no. 1, pp. 18-21, Oct 2015.
- [3] B.V. Dasarathy, Nearest Neighbor Pattern Classification Techniques. California: IEEE Computer Society Press, 1991.

- [4] G. Shakhnarovich, T. Darrel and P. Indyk, Nearest Neighbor Methods in Learning and Vision, MIT Press, 2006.
- [5] H. Jiawei and K. Micheline, Data Mining Concepts and Techniques, 2nd ed. USA: Morgan Kaufmann, 2006.
- [6] J.T. Barkoulas and C.F. Baum, "Long-term dependence in stock returns", Economics Letter, pp.253-259, Oct. 1996.
- [7] L. Khaidem, S. Saha and S.R. Dey, "Predicting the direction of stock market prices using random forest", Applied Mathematical Finance, vol. 0, no. 0, pp. 1-20, May. 2015.
- [8] P. Sanguansat, Machine Learning. Thailand: Panyapiwat Institute of Management, 2015
- [9] P. Sutheebanjard and W. Premchaiswadi. "Stock Exchange of Thailand Index prediction using Back Propagation Neural Network", Second International Conference on Computer and Network Technology, pp. 377-380, Apr. 2010.
- [10] S. Chaigusin, C. Chirathamjaree and J. Clayden. "The Use of Neural Networks in the Prediction of the Stock Exchange of Thailand (SET) Index", International Conference on Computational Intelligence for Modeling Control & Automation, pp.670-673, Dec. 2008.
- [11] S. Nootyaskool, M. Attasiripanya and W. Prasitichaiikul. "Stock Forecasting with Candlestick Chart and Hidden Markov Models", National Conference on Computing and Information Technology, 2014.
- [12] W. Huang, Y. Nakamori and S.Y. Wang, "Forecasting stock market movement direction with support vector machine", Computers & Operations Research, vol. 32, no.10, pp. 2513-2522, Oct. 2005.
- [13] W. Qiang, C. Yuehui and L. Zhen, "Ensemble model of Intelligent Paradigms for stock Market Forecasting", Knowledge Discovery and Data Mining, pp. 205-208, Jan. 2008.
- [14] Z. H. Zhou, Ensemble Methods Foundations and Algorithms, New York: CRC Press, 2012.