ปัจจัยที่มีผลต่อการผันแปรของหุ้นกลุ่มเทคโนโลยีโดยใช้เทคนิคการทำเหมืองข้อมูล

FACTORS AFFECTING THE VARIATION OF TECHNOLOGY STOCKS USING BY DATA MINING

ยศสยา แสงหิรัญ[1] และ สมชาย เล็กเจริญ[2]

[1] นักศึกษาระดับปริญญาโท หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ
วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต, yossaya.s56@rsu.ac.th
[2] อาจารย์ประจำ วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต, somchai.l@rsu.ac.th

---

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลกระทบต่อการผันแปรของราคาหุ้นกลุ่มเทคโนโลยี ราคาปิดรายเดือน และเพื่อเปรียบเทียบประสิทธิภาพของโมเดล โดยนำเทคนิคการทำเหมืองข้อมูลมาใช้ ได้แก่ อัลกอริทึมต้นไม้ตัดสินใจ (J48) อัลกอริทึมการจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest) อัลกอริทึมการแยกประเภทแบบการหาเพื่อนบ้านใกล้ที่สุด (K- Nearest Neighbor: KNN) อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network) เพื่อหารูปแบบการพยากรณ์ที่เหมาะสมสำหรับการผันแปรของราคาหุ้นกลุ่มเทคโนโลยีราคาปิดรายเดือน และศึกษาปัจจัยที่มีผลต่อการผันแปรของราคาหุ้นกลุ่มเทคโนโลยีราคาปิดรายเดือน ใช้การวิเคราะห์ด้วยสถิติถดถอยพหุคูณ ในการศึกษาครั้งนี้ใช้ข้อมูลจากตลาดหลักทรัพย์แห่งประเทศไทย เป็นระยะเวลาย้อนหลัง 5 ปี ตั้งแต่ปี 2556-2560 จำนวนหุ้นทั้งหมด 36 หุ้น ซึ่งสามารถพิจารณาเป็นปัจจัยที่มีผลต่อการผันแปรมี 10 องค์ประกอบดังต่อไปนี้ เวลาปิดตลาด (Close), มูลค่าหลักทรัพย์ตามตลาด (Market Cap), อัตราส่วนราคาตลาดต่อกำไรสุทธิ (P/E), อัตราส่วนราคาตลาดต่อมูลค่าตามบัญชี (P/BV), มูลค่าตามบัญชีต่อหุ้น Book Value Per Share (BVPS), อัตราผลตอบแทนจากเงินปันผล Dividend Yield (DivY), อัตราการหมุนเวียนการซื้อขาย Turnover Ratio, อัตราผลตอบแทนผู้ถือหุ้น Return On Equity (ROE) , ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA) และสินทรัพย์สุทธิ (Tota Asset) โดยใช้การวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression Analysis) พบว่าค่าความแปรปรวนตัวแปรของหุ้นกลุ่มเทคโนโลยีที่มีผลต่อการเปลี่ยนแปลงของราคาปิดของหุ้นแต่ละเดือนมากที่สุดคือ มูลค่าตามบัญชีต่อหุ้น Book Value Per Share (BVPS) รองลงมาคือ ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA) และสุดท้ายคือ อัตราผลตอบแทนผู้ถือหุ้น Return On Equity (ROE) ตามลำดับความสำคัญ และเมื่อทดสอบแบบจำลองพบว่าการแยกประเภทแบบการหาเพื่อนบ้านใกล้ที่สุด (K- Nearest Neighbor: KNN) มีประสิทธิภาพดีที่สุด ซึ่งมีค่าความแม่นยำอยู่ที่ร้อยละ 100 และค่าความคลาดเคลื่อนเฉลี่ยอยู่ที่ร้อยละ 0.00 รองลงมา การจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest) มีค่าความแม่นยำอยู่ที่ร้อยละ 100 และมีค่าความคลาดเคลื่อนอยู่ที่ร้อยละ 0.14 โครงข่ายประสาทเทียม (Neural Network) มีค่าความแม่นยำอยู่ที่ร้อยละ 81.11 และมีค่าความคลาดเคลื่อนอยู่ที่ร้อยละ 0.24 และต้นไม้ตัดสินใจ J48 มีค่าความแม่นยำอยู่ที่ 76.66% และมีค่าความคลาดเคลื่อน อยู่ที่ 0.35 ตามลำดับประสิทธิภาพ ดังนั้นผลวิจัยครั้งนี้มีประโยชน์สำหรับผู้ที่จะลงทุนหุ้นกลุ่มเทคโนโลยีต้องคำนึงถึงมูลค่าตามบัญชีต่อหุ้น Book Value Per Share (BVPS) รองลงมาคือ ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA) และสุดท้ายคือ อัตราผลตอบแทนผู้ถือหุ้น Return On Equity (ROE)

คำสำคัญ: การทำเหมืองข้อมูล, การวิเคราะห์การถดถอยเชิงพหุคูณ, หุ้นกลุ่มเทคโนโลยี

ABSTRACT

The objectives of this study were to study factors affecting to variation of the monthly closing price of Technology share and to compare efficiency of model by data mining technic such as Decision Tree (J48), Random Forest, K- Nearest Neighbor: KNN and Neural Network in order to find the right forecasting model for the monthly closing price of Technology share and to study factors affecting to variation of the monthly closing price of Technology share by multiple regression analysis. Data used were collected from The Securities Exchange of Thailand for the past 5 years from 2013-2017, the total number of shares was 36 shares. The factors that affected the variation were the following 10 elements; Close, Market Cap, P/E, P/BV, Book Value Per Share (BVPS), Dividend Yield (DivY), Turnover Ratio, Return On Equity (ROE) , Return on Assets (ROA) and Tota Asset. The data was analyzed by Multiple Linear Regression Analysis. It was found that variation of the Technology share which affected to the change in closing price of each month, the most was Book Value Per Share (BVPS), secondary was Return on Assets (ROA) and Return On Equity (ROE), respectively. The model test was found that classification by K- Nearest Neighbor: KNN had been the most effective, the accuracy value was 100% and the average deviation was 0.00% following by Random Forest, the accuracy value was 100% and the average deviation was 0.14 % besides Neural Network, the accuracy value was 81.11% and the average deviation was 0.24 and Decision Tree (J48), the accuracy value was 76.66% and the average deviation was 0.35%, respectively as efficiency. The result of this study has benefit for investors who will invest in Technology share must aware of Book Value Per Share (BVPS), Return on Assets (ROA) and Return On Equity (ROE), respectively.

Keywords: Data mining, Multiple Linear Regression Analysis, Technology stock

1. บทนำ

เนื่องด้วยในภาวะปัจจุบันมีเครื่องมือทางการเงินหลากหลายประเภทเกิดขึ้นเพื่อเป็นทางเลือกให้ กับผู้ลงทุน ผู้ลงทุนส่วนใหญ่มักเน้นไปที่การออม การฝากเงินไว้กับธนาคารพาณิชย์หรือสถาบันการเงินต่างๆ ยิ่งผู้ลงทุนฝากมาก ดอกเบี้ยที่ได้รับก็ยิ่งมีมาก ซึ่งเป็นวิธีที่ปลอดภัยและมีความเสี่ยงน้อย แต่อาจจะต้องใช้เป็นเวลานานกว่าจะได้รับดอกเบี้ยตามที่ต้องการ ในปัจจุบันการลงทุนนั้นได้แพร่หลายมากยิ่งขึ้น ทั้งการลงทุนในอสังหาริมทรัพย์ การลงทุนตราสารทุนหรือหุ้น ตราสารหนี้หรือ หุ้นกู้ การออมและการเกรงกำไรทอง ทำให้มีนักลงทุนสนใจที่จะลงทุนเพิ่มขึ้นเป็นจำนวนมาก ถึงแม้ความเสี่ยงจะยังมีเหมือนเดิมก็ตาม ในการวิเคราะห์และการตัดสินใจเลือกซื้อหุ้น ผู้ลงทุนต้องอาศัยข้อมูลจำนวนมากเพื่อการวิเคราะห์ประกอบกับการตัดสินใจที่รวดเร็ว การลงทุนจึงจะมีประสิทธิภาพ ความเสี่ยงต่ำ และผลตอบแทนสูง เพื่อเพิ่มประสิทธิภาพในการตัดสินใจเลือกลงทุน (ฉัตรแพร พูนเอียด, 2558)

นักลงทุนส่วนใหญ่มักหาข่าวสารและข้อมูลที่มีผลต่อราคาหุ้นมาใช้เป็นตัวแปรในการวิเคราะห์เพื่อหาปัจจัยและโอกาส ที่จะได้ผลตอบแทนจากการลงทุนมากที่สุด โดยการวิเคราะห์มี 2 รูปแบบที่ คือ การวิเคราะห์ปัจจัยพื้นฐานและการวิเคราะห์ปัจจัยทางเทคนิค

การวิเคราะห์ปัจจัยพื้นฐาน (Fundamental analysis) หมายถึง การประเมินผลการดำเนินงาน รายได้ กำไรสุทธิ เพื่อคาดการณ์อนาคต โดยใช้การพิจารณาจากปัจจัยที่มีผลกระทบต่อผลการดำเนินงานในอนาคต อาทิ ภาวะ

เศรษฐกิจ การเงิน การเมือง ภาวะอุตสาหกรรมที่เกี่ยวข้อง แนวทางการบริหารงานของผู้ออกผลิตภัณฑ์ในตลาดทุน รวมถึงความเสี่ยงต่าง ๆ ทางธุรกิจ (ศิริวรรณ พิมเสน, 2558)

การวิเคราะห์ปัจจัยทางเทคนิค (Technical Analysis) หมายถึง การวิเคราะห์การเคลื่อนไหวของราคาหลักทรัพย์ในอดีต โดยใช้หลักสถิติ เพื่อคาดการณ์ราคาที่น่าจะเป็นในอนาคต โดยเชื่อราคาปัจจุบันสะท้อนข่าวสารในด้านต่างๆ ทั้งหมดแล้ว การเคลื่อนไหวของราคมมีลักษณะเป็นแนวโน้ม และมีการเคลื่อนไหวของราคาที่คล้ายคลึงกับในอดีตรูปแบบของกราฟที่ใช้ในการวิเคราะห์ (ศิริวรรณ พิมเสน, 2558)

การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล (ศจี วานิช, 2558)

การตัดสินใจ (Decision Making) หมายถึง กระบวนการในการตัดสินใจที่กระทำสิ่งใดสิ่งหนึ่งจากทางเลือกต่างๆที่มีอยู่ ซึ่งผู้บริโภคมักจะต้องตัดสินใจในทางเลือกต่างๆของสินค้าและบริการอยู่เสมอ โดยที่เขาจะเลือกสินค้าหรือบริการตามข้อมูลและข้อจำกัดของสถานการณ์การตัดสินใจจึงเป็นกระบวนการที่สำคัญและอยู่ภายในจิตใจของผู้บริโภค ขั้นตอนเพื่อนำไปสู่การตัดสินใจซื้อสินค้าหรือบริการ มีทั้งหมด 5 ขั้นตอน การตระหนักถึงความต้องการ การค้นหาข้อมูล การประเมินทางเลือก การตัดสินใจซื้อพฤติกรรมภายหลังซื้อ (คฑาเทพ นิติสิริ, 2558)

ต้นไม้ตัดสินใจ (Decision Tree) หมายถึง การนำข้อมูลมาสร้างแบบจำลองเพื่อการทำนายหรือการจำแนกออกเป็นประเภทต่างๆ โดยมีโครงสร้างในลักษณะที่เป็นต้นไม้ โดยในการจำแนกดังกล่าวจำเป็นจะต้องมีการเรียนรู้แบบ Supervised Learning คือการมีข้อมูลสอนเพื่อจำแนกโครงสร้างการตัดสินใจหรือการทำนายไว้ล่วงหน้าเพื่อใช้ในการทำนายหรือจำแนกข้อมูลที่เข้ามาใหม่ (วิภาวรรณ บัวทอง, 2557)

Decision Tree Random Forest หมายถึง เทคนิคที่คล้ายๆ กับ Bagging แต่แทนที่จะสุ่มข้อมูลอย่างเดียวก็ทำการสุ่มเลือกแอตทริบิวต์ (ฟีเจอร์) ต่างๆ ออกมาเป็นหลายๆ ชุด และสร้างโมเดลด้วยเทคนิค Decision Tree หลายๆ ต้น แล้วนำผลที่ได้แต่ละโมเดลมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย (ศริญณา มาปลูก, 2559)

การหาเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) หมายถึง วิธีการจำแนกประเภทข้อมูลวิธีการหนึ่งโดยจัดเป็นวิธีการจำแนกประเภทข้อมูลแบบมีผู้ฝึกสอน (Supervised Learning) หรือการที่ทราบคำตอบของข้อมูลอยู่ก่อนแล้ว จากนั้นใช้โมเดลในการจำแนกประเภทข้อมูลจากข้อมูลฝึกที่ทราบคำตอบโดยการตรวจสอบจำนวนบางจำนวน (“K” ใน K-nearest neighbor) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่างๆสำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด (พงศกร ธีรรัศมี, 2558)

โครงข่ายประสาทเทียม (Neural Networks) หมายถึง ตัวแบบทางคณิตศาสตร์ที่จำลองการทำงานของเซลประสาทของสิ่งมีชีวิตซึ่งประกอบขึ้นจากแบบจำลองของโหนดจำนวนมากมายเชื่อมต่อกันเป็นโครงข่าย บางโหนดเชื่อมต่อกับสิ่งแวดล้อมภายนอกและสมารถทำหน้าที่เป็นหน่วย Input หรือ Output ได้ แนวคิดเริ่มต้นของเทคนิคนี้

ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (Neurons) และ จุดประสานประสาท (Synapses) (เกรียงไกร ราชกิจ, 2561)

จากโมเดลข้างต้นต้องได้รับการประเมินเพื่อให้มั่นใจว่าตรงกับหลักเกณฑ์ความสำเร็จของการทำเหมืองข้อมูลและผ่านหลักเกณฑ์การทดสอบ เทคนิคการประเมินทั้งหมดนี้มาจากผลลัพธ์ของขั้นตอนการสร้างตัวแบบ

		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Accuracy คือความถูกต้องจากการทำนายถูกทั้ง แบบ Positive และ Negative โดยมีสูตร

$$\text{Accuracy} = \frac{(TP+TN)}{TP+FN+FP+TN} \quad X\ 100$$

ค่าความคลาดเคลื่อนกำลังสองของค่าเฉลี่ย (Mean squared error: MSE)

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยใช้หลักการเดียวกับการหาค่าความแปรปรวนในทางสถิติ การวัดค่าความคลาดเคลื่อนด้วนวิธีนี้จะได้ค่าความคลาดเคลื่อนที่สูง เนื่องจากเป็นการนำความคลาดเคลื่อน ณ เวลาใดๆมายกกำลังสองก่อนที่จะหาผลรวม แล้วจึงนำมาหาค่าเฉลี่ยอีกครั้งหนึ่ง นั้นคือ ค่า MSE ยิ่งน้อย หมายถึง การพยากรณ์ยิ่งแม่นยำ โดยมีสูตร

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n}$$

โดยที่ y_i แทน ค่าจริง; $\hat{y_i}$ แทนค่าพยากรณ์

การวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression Analysis) ซึ่งเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันทำนายหรือพยากรณ์หรืออธิบายการผันแปรของตัวแปรตามได้ เขียนเป็นสมการดังนี้

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

x_i คือ ค่าของตัวแปรอิสระแต่ละตัว

Y คือ ค่าของตัวแปรตาม

β_0 คือ ค่าคงที่ (Constant) ของสมการถดถอย

β_i คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระ x_i แต่ละตัว (จะใช้สัญลักษณ์ bi สำหรับค่าที่ได้จากตัวอย่าง และ สำหรับค่าประมาณหรือตัวทำนาย) β_i หรือ bi จะแสดงอัตราการเปลี่ยนแปลงของค่า x_i ต่อค่า Y ดังนี้ คือ ถ้าค่า x_i เปลี่ยนไป 1 หน่วย จะทำให้ค่า Y เปลี่ยนไป bi หน่วย

ε คือ ค่าความคลาดเคลื่อน (Error or Residual) (สุทิน ชนะบุญ, 2560)

จากการสำรวจในปี 2559 ปรากฏว่าหุ้นกลุ่มเทคโนโลยีมีค่า PE-P/BV ต่ำแต่มีผลตอบแทนสูงกว่าตลาด โดยนำข้อมูลจากตลาดหลักทรัพย์แห่งประเทศไทย ดังนั้นผู้วิจัยมีข้อสนใจ และได้นำเอาข้อมูลหรือตัวแปรที่มีเกี่ยวข้องกับความผันแปรของหุ้นมาวิเคราะห์ ผ่านอัลกอริทึมต่างๆ ด้วยเทคนิคการทำเหมืองข้อมูล (Data mining) เพื่อเปรียบเทียบและหาความแม่นยำของอัลกอริทึมและนำผลลัพธ์มาสนับสนุนการตัดสินใจลงทุนหุ้นในกลุ่มเทคโนโลยี ในปัจจุบันการทำเหมืองข้อมูลได้นำเอาไปประยุกต์ใช้กับงานในหลากหลายด้านการวิเคราะห์ในครั้งนี้เป็นการเจาะจงวิเคราะห์หุ้นในกลุ่มเทคโนโลยี (TECH) ซึ่งจะแบ่งแยกย่อยได้ 2 หมวด ได้แก่ หมวดชิ้นส่วนอิเล็กทรอนิกส์ (ETRON) และหมวดเทคโนโลยีสารสนเทศและการสื่อสาร (ICT) โดยนำข้อมูลจาก SETSMART (SET Market Analysis and Reporting Tool) เป็นบริการระบบข้อมูลตลาดหลักทรัพย์ฉบับออนไลน์ ผลิตโดยตลาดหลักทรัพย์แห่งประเทศไทย มาวิเคราะห์การผันแปรของราคาหุ้นเพื่อช่วยสนับสนุนตัดสินใจในการลงทุน

2. วัตถุประสงค์การวิจัย

1. เพื่อศึกษาปัจจัยที่มีอิทธิผลต่อการผันแปรของหุ้นกลุ่มเทคโนโลยี โดยใช้เทคนิคการทำเหมืองข้อมูล
2. เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองของเทคนิคการทำเหมืองข้อมูล

3. การดำเนินการวิจัย

การวิจัยครั้งนี้มีขั้นตอนพอสรุปได้ดังนี้

3.1 การเตรียมข้อมูล

รวบรวมข้อมูลหุ้นกลุ่มเทคโนโลยี (TECH) หมวดหมวดชิ้นส่วนอิเล็กทรอนิกส์ (ETRON) และหมวดเทคโนโลยีสารสนเทศและการสื่อสาร (ICT) จาก www.setsmart.com และเก็บบันทึกค่าตัวแปรทางการเงินที่สำคัญดังนี้ แนวโน้มราคาหุ้น, มูลค่าหลักทรัพย์ตามราคาตลาด, อัตราส่วนราคาตลาดต่อกำไรสุทธิ, อัตราส่วนราคาตลาดต่อมูลค่าตามบัญชี, มูลค่าตามบัญชีต่อหุ้น, อัตราผลตอบแทนจากเงินปันผล, อัตราการหมุนเวียนการซื้อขาย, อัตราผลตอบแทนผู้ถือหุ้น, ผลตอบแทนต่อสินทรัพย์, สินทรัพย์สุทธิและราคา ณ เวลาปิดตลาด ตั้งแต่ปี พ.ศ. 2556 ถึง พ.ศ. 2560 เป็นระยะเวลา 5 ปี ทั้งหมด 36 หุ้น ลงในโปรแกรม Microsoft Excel และบันทึกนามสกุลไฟล์เป็น .CSV

3.2 การหาค่าปัจจัยที่มีผลต่อการเปลี่ยนแปลงของราคาหุ้นกลุ่มเทคโนโลยี

นำข้อมูลหุ้นกลุ่มเทคโนโลยีที่เก็บบันทึกไว้แบบไฟล์ .CSV นำเข้ามาใส่ในโปรแกรมวิเคราะห์ทางสถิติข้อมูล และใช้การวิเคราะห์ข้อมูลด้วยวิธีแบบสมการถดถอยพหุคูณ (Multiple Regression Analysis) โดยนักวิจัยได้ใช้ ค่าราคา ณ เวลาตลาดปิด ตั้งเป็นตัวแปรตาม (Dependent Variable) และนำค่า มูลค่าหลักทรัพย์ตามราคาตลาด, อัตราส่วนราคา

ตลาดต่อกำไรสุทธิ, อัตราส่วนราคาตลาดต่อมูลค่าตามบัญชี, มูลค่าตามบัญชีต่อหุ้น, อัตราผลตอบแทนจากเงินปันผล, อัตราการหมุนเวียนการซื้อขา, อัตราผลตอบแทนผู้ถือหุ้น, ผลตอบแทนต่อสินทรัพย์ และสินทรัพย์สุทธิ มาใส่เป็นตัวแปรอิสระ (Independent Variable) และเก็บผลลัพธ์จากการคำนวณเพื่อมาสรุปผลของตัวแปรที่มีผลต่อการเปลี่ยนแปลงของหุ้นกลุ่มเทคโนโลยี

3.3 การทดสอบ

นำข้อมูลหุ้นกลุ่มเทคโนโลยีที่เตรียมไว้ มาประมวลผลแบบการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลโดยใช้โปรแกรมวิเคราะห์ข้อมูล Weka เพื่อจัดแบ่งประเภทของข้อมูล เช่น ต้นไม้การตัดสินใจ (Decision Tree) โดยอัลกอริทึมที่ใช้ทดสอบมีดังนี้ Decision Tree J48, Decision Tree RandomForest, K-Nearest Neighbor (KNN), Neural Network ทำการหาค่าความแม่นยำ (Accuracy) และหาค่าความคลาดเคลื่อน (Mean squared error: MSE) ของอัลกอริทึมนั่นเพื่อทำการเปรียบเทียบ และหาแบบจำลองที่ดีที่สุดจากวิธีการทดสอบหาค่าพยากรณ์ความถูกต้องของข้อมูลว่ามีความถูกต้องมากน้อยเพียงใดโดยคิดเป็นค่าร้อยละ

หาปัจจัยที่มีผลต่อการเปลี่ยนแปลงของราคาหุ้นกลุ่มเทคโนโลยี ใช้สถิติการวิเคราะห์การถดถอยเชิงพหุคูณ ในโดยใช้โปรแกรมสำเร็จรูปทางสถิติ

4. ผลการวิจัย

ผลจากการทดสอบแบ่งออกเป็น 2 ส่วนตามวัตถุประสงค์ดังนี้

1. ผลจากการศึกษาหาปัจจัย ผู้วิจัยได้นำผลจากการการวิเคราะห์ด้วยเทคนิคการวิเคราะห์การถดถอยเชิงเส้นตรงพหุคูณ (Multiple Regression Analysis) ที่มีผลต่อการเปลี่ยนแปลงของราคาหุ้นกลุ่มเทคโนโลยี ได้ผลตามตารางที่ 1 ตารางที่ 2 และตารางที่ 3

ตารางที่ 1 โมเดลสรุปผล

R	R Square	Adjusted R Square	Std. Error of the Estimate
.96	0.93	0.93	14.46

จากตารางที่ 1 พบว่า R Square = 0.93 หรือ 93% แสดงว่า โมเดลนี้ สามารถพยากรณ์ได้ถูกต้อง 93%

ตารางที่ 2 การวิเคราะห์ความแปรผันของหุ้นกลุ่มเทคโนโลยี

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	479407.51	9	53267.50	254.54	.000[b]
Residual	35575.63	170	209.26		
Total	514983.14	179			

P value < 0.05

จากตารางที่ 2 พบว่าตัวแปรที่ใช้การพยากรณ์นั้นสามารถพยากรณ์ได้ อย่างมีนัยสำคัญทางสถิติที่ ระดับ 0.05 แสดงว่าการพยากรณ์ในครั้งนี้สามารถนำตัวแปรต่างๆมาพยากรณ์การเปลี่ยนแปลงราคาปิดของหุ้นกลุ่มเทคโนโลยีแต่ละเดือน

การทดสอบวิเคราะห์หาค่าสัมประสิทธิ์ โดยการนำข้อมูล ราคา ณ เวลาปิดตลาด (Close), มูลค่าหลักทรัพย์ตามตลาด (Market Cap), อัตราส่วนราคาตลาดต่อกำไรสุทธิ (P/E), อัตราส่วนราคาตลาดต่อมูลค่าตามบัญชี (P/BV), มูลค่าตามบัญชีต่อหุ้น Book Value per Share (BVPS), อัตราผลตอบแทนจากเงินปันผล Dividend Yield (DivY), อัตราการหมุนเวียนการซื้อขาย Turnover Ratio, อัตราผลตอบแทนผู้ถือหุ้น Return On Equity (ROE) , ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA)และสินทรัพย์สุทธิ (TotaAsset) วิเคราะห์โดยใช้การวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression Analysis) เขียนเป็นสมการดังนี้

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

และเมื่อนำมาใส่ค่าตัวแปรจะได้สมการดังนี้ Y = 6.54 + (0.00*MC) + (-0.01*P/E) + (0.00*P/BV) + (0.85*BVPS) + (0.00*DivY) + (-0.01*Turnover Ratio) + (0.07*ROE) + (0.10*ROA) + (0.00*TA)

จากผลการวิเคราะห์ค่าความแปรปรวนตัวแปรของหุ้นกลุ่มเทคโนโลยีที่มีผลต่อการเปลี่ยนแปลงของราคาปิดของหุ้นแต่ละเดือนมากที่สุดคือ มูลค่าตามบัญชีต่อหุ้น Book Value per Share (BVPS) รองลงมาคือ ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA) และสุดท้ายคือ อัตราผลตอบแทนผู้ถือหุ้น Return On Equity (ROE)

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของแบบจำลองโมเดลแต่ละชนิด

Algorithm	Accuracy	Mean squared Error
KNN	100.00	0.00
Random Forest	100.00	0.14
Neural Network	81.11	0.24
J48	76.66	0.35

จากตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองของโมเดล ทั้งหมด 4 แบบ แบบจำลองของโมเดลต้นไม้ตัดสินใจ (Decision Tree) J48, Random Forest, KNN และ Neural Network ผ่านโปแกรมทำเหมืองข้อมูล Weka ใช้ Training set โดยกำหนดค่า Cross-validation Folds = 10 ได้ผลการทดสอบดังตารางที่ 3 โดยแบบจำลองโมเดลที่ทำการวิเคราะห์ทั้งหมด ได้ผลลัพธ์ที่ดีที่สุด คือ อัลกอริทึมชนิด KNN ซึ่งมีค่าความแม่นยำอยู่ที่ 100% และมีค่าความคลาดเคลื่อน อยู่ที่ 0.00 รองลงมาเป็น อัลกอริทึมชนิด Random Forest มีค่าความแม่นยำอยู่ที่ 100% และมีค่าความคลาดเคลื่อนอยู่ที่ 0.14 อัลกอริทึมชนิด Neural Network ซึ่งมีค่าความแม่นยำอยู่ที่ 81.11% และมีค่าความคลาดเคลื่อน อยู่ที่ 0.24 และอัลกอริทึมชนิด J48 มีค่าความแม่นยำอยู่ที่ 76.66% และมีค่าความคลาดเคลื่อน อยู่ที่ 0.35 สรุปว่าการพยากรณ์ข้อมูลนี้ ใช้โมเดล KNN สามารถพยากรณ์ได้แม่นยำ 100% และมีค่าผิดพลาดที่ 0.00

5. บทสรุปและข้อเสนอแนะ

ผลการวิจัยเพื่อหาปัจจัยที่มีผลกระทบต่อการแปรผันของราคาหุ้นกลุ่มเทคโนโลยีและการเปรียบเทียบของแบบจำลองโมเดลด้วยวิธีการทำเหมืองข้อมูลของหุ้นกลุ่มเทคโนโลยี ในตลาดหลักทรัพย์แห่งประเทศไทยทั้งหมด 36 บริษัท โดยใช้ข้อมูลย้อนหลัง 5 ปี นับจากปี 2556 – 2560 มีประเด็นอภิปรายดังนี้

1. จากการศึกษาและการทดสอบเกี่ยวกับปัจจัยที่ส่งผลต่อการเปลี่ยนแปลงของราคาหุ้นกลุ่มเทคโนโลยีของตลาดหลักทรัพย์แห่งประเทศไทย โดยการวิเคราะห์แบบการถดถอยพหุคูณ (Multiple Regression Analysis) ปรากฏว่า

ปัจจัยที่มีผลกระทบมากที่สุด คือ มูลค่าตามบัญชีต่อหุ้น Book Value per Share (BVPS) รองลงมาคือ ผลตอบแทนต่อสินทรัพย์ Return on Assets (ROA) และสุดท้ายคือ อัตราผลตอบแทนผู้ถือหุ้น Return on Equity (ROE)

2. ผลการทดสอบจากการเปรียบเทียบประสิทธิภาพและความถูกต้องของแบบจำลองโมเดล แบบจำลองที่ดีที่สุด คือ แบบจำลองการแยกประเภทแบบการหาเพื่อนบ้านใกล้ที่สุด (K- Nearest Neighbor: KNN) ซึ่งมีค่าความแม่นยำอยู่ที่ 100% และมีค่าความคลาดเคลื่อน อยู่ที่ 0.00 (จากค่าในตารางที่ 3)

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. นักวิจัยและผู้สนใจด้านงานวิจัยสามารถนำผลการวิจัยนี้ไปพัฒนาและประยุกต์ให้งานวิจัยของนักวิจัยมีประสิทธิภาพ

2. ในการใช้งานจริงงานวิจัยชิ้นนี้เป็นเครื่องมือช่วยสนับสนุนการตัดสินใจในการลงทุนหุ้นของกลุ่มเทคโนโลยีที่อ้างอิงจากการวิเคราะห์ข้อมูลจากปี พ.ศ. 2556 – 2560 และผลลัพธ์จะมีการเปลี่ยนแปลงเสมอเมื่อผู้ใช้งานเพิ่มตัวแปรและข้อมูลใหม่เข้าไปในการวิเคราะห์

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1.ในการศึกษาครั้งถัดไปควรเพิ่มจำนวนตัวแปรและความถี่ของข้อมูลที่ใช้ในการวิเคราะห์เพื่อให้ได้ผลลัพธ์ที่ละเอียดและแม่นยำมากกว่าเดิม

2. การวิเคราะห์หุ้นกลุ่มเทคโนโลยีโดยการใช้เทคนิคเหมืองข้อมูลครั้งนี้ได้ใช้แบบจำลองโมเดลต้นไม้ตัดสินใจเพียงชนิดเดียว เพื่อความถูกต้องและแม่นยำที่สุดควรทดสอบโดยใช้แบบจำลองที่หลากหลายแล้วนำมาเปรียบเทียบกัน

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก ผศ.ดร.สมชาย เล็กเจริญ อาจารย์ที่ปรึกษาการค้นคว้าอิสระซึ่งได้ให้ความรู้ การชี้แนะแนวทางการศึกษา ตรวจทานและแก้ไขข้อบกพร่องในงาน ตลอดจนการให้คำปรึกษาซึ่งเป็นประโยชน์ในการวิจัยจนงานวิจัยครั้งนี้มีความสมบูรณ์ครบถ้วนสำเร็จได้ด้วยดี และระบบข้อมูลจาก SETSMART ที่เก็บรวบรวมข้อมูลทั้งหมดของตลาดหลักทรัพย์แห่งประเทศไทยทั้งหมดบนอินเตอร์เน็ตผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูง มาไว้ ณ โอกาสนี้

เอกสารอ้างอิง

เกียงไกร ราชกิจ. (2561). *โครงข่ายประสาทเทียม หรือ ข่ายงานประสาทเทียม (artificial neural network).* สืบค้น 20 พฤษภาคม 2561 จาก www.erp.mju.ac.th/acticleDetail.aspx?qid=764

คฑาเทพ นิติสิริ. (2558). *ปัจจัยที่มีอิทธิพลต่อแนวโน้มการตัดสินใจใช้บริการสินเชื่อแบบ Reverse Mortgage*. หลักสูตรบริหารธุรกิจมหาบัณฑิต, มหาวิทยาลัยกรุงเทพ.

ฉัตรแพร พูนเอียด. (2558). *ปัจจัยและพฤติกรรมในการตัดสินใจลงทุนในกองทุนรวมหุ้นระยะยาวของนักลงทุนในเขตกรุงเทพมหานคร*. วิทยานิพนธ์เศรษฐศาสตร์รมหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์.

พงศกร ธีรรัศมี. (2558). *วิธีการหาค่า เค ที่เหมาะสมในการจำแนกแบบเคเนียร์เรสเนเบอร์กับข้อมูลทางการแพทย์.* ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี.

วิภาวรรณ บัวทอง. (2557*). Chapter 5 Data Mining for Classification*. สืบค้น 22 พฤษภาคม 2561 จาก https://wipawanblog.files.wordpress.com/2014/06/chapter-5- classification.pdf

ศจี วานิช. (2558). *Data mining เหมืองข้อมูล*. สืบค้น 2 มีนาคม 2561, จาก http://sajeegm301.blogspot.com/2015/11/data-mining.html

ศริญฌา มาปลูก. (2559). *การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining โดยซอฟต์แวร์ RapidMiner Studio 6*. สืบค้น 21 พฤษภาคม 2561, จาก http://www.erp.mju.ac.th/acticleDetail.aspx?qid=579

ศิริวรรณ พิมเสน. (2558). *การวิเคราะห์หลักทรัพย์โดยวิธีทางเทคนิค*. สืบค้น 15 พฤษภาคม 2561, จาก https://prezi.com/4njpjx-7bm_k/presentation/

สุทิน ชนะบุญ. (2560). *สถิติและการวิเคราะห์ข้อมูลในงานวิจัยด้านสุขภาพเบื้องต้น*, สำนักงานสาธารณะสุขจังหวัดขอนแก่น.