

**การทำเหมืองข้อมูลจากงบการเงินเพื่อทำนายราคาตลาด
สำหรับหุ้นในกลุ่ม SET50 กรณีศึกษาตลาดหลักทรัพย์แห่งประเทศไทย**
**Data Mining of Financial Statement for Predicting Common Stock Prices for SET50:
A Case Study of Thailand**

ฐะปะนีย์ ตีรรัตนภรณ์* และวชรรักษ์ พรหมมา²

Tapanee Treeratanaporn* and Watcharawat Promma²

Received : October 4, 2021 Revised : May 31, 2022 Accepted : June 10, 2022

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ในการนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์เพื่อหาปัจจัยที่มีผลต่อราคาตลาดของหุ้นสามัญกลุ่ม SET50 ในตลาดหลักทรัพย์แห่งประเทศไทยและเพื่อทำนายราคาตลาดของหุ้นสำหรับการลงทุนในอนาคตประชากรในงานวิจัยนี้คือ บริษัทที่จดทะเบียนในกลุ่ม SET 50 เป็นข้อมูลงบการเงินแบบฟอร์ม 56-1 ตั้งแต่ปี พ.ศ. 2559 ถึงปี พ.ศ. 2563 ตัวแปรต้นที่ศึกษา ได้แก่ ทุนจดทะเบียน จำนวนหุ้นสามัญ ส่วนของผู้ถือหุ้น รายได้ กำไรสุทธิ กำไรต่อส่วนของผู้ถือหุ้นสามัญ และการจ่ายเงินปันผล ขั้นตอนการดำเนินงานเริ่มจากเก็บรวบรวมข้อมูลทั้งห้าปี นำข้อมูลที่ได้มาทำความสะอาด ปรับแต่ง และวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูลในสามรูปแบบ คือ Generalized Regression Model, Decision Tree และ Support Vector Machine โดยใช้เครื่องมือโปรแกรม RapidMiner Go ผลการวิจัยพบว่า เมื่อใช้ทั้งสามเทคนิคที่แตกต่างกันออกไป ผลลัพธ์ที่ได้จากความสัมพันธ์แต่ละตัวแปรก็มีผลที่ต่างกัน อย่างไรก็ตามเมื่อพิจารณาภาพรวมทั้งสามเทคนิคพบว่า ตัวแปรต้นทั้ง 7 ตัวที่เลือกมาศึกษา ล้วนมีผลต่อราคาตลาดของหุ้นสามัญในกลุ่ม SET 50 ทั้งสิ้น ผลการทำนายพิจารณาจากค่า RMSE (Root Mean Square Error) เป็นสำคัญ ซึ่งถ้าค่า RMSE มีค่าน้อยแสดงว่ามีความผิดพลาดน้อย ค่า RMSE ของเทคนิค Decision Tree มีค่าเท่ากับ 28.19 Support Vector Machine เท่ากับ 35.05 และ Generalized Regression Model เท่ากับ 40.78 ตามลำดับ ดังนั้น Decision Tree จึงให้ผลการทำนายที่มีความถูกต้องมากที่สุดในงานสามเทคนิคนี้ สำหรับการแสดงผลของ Decision Tree จะแสดงผลในรูปแบบแผนภูมิต้นไม้ ส่วน Support Vector Machine แสดงผลเป็นค่าน้ำหนัก และ Generalized Regression Model แสดงผลเป็นค่าสัมประสิทธิ์

คำสำคัญ : การทำเหมืองข้อมูล การทำนาย ราคาหุ้น งบการเงิน

* ผู้ช่วยศาสตราจารย์ ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* Assistant Professor, Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok

² อาจารย์ประจำ สาขาวิชาบัญชี สำนักวิชาการจัดการ มหาวิทยาลัยวลัยลักษณ์

² Lecturer, Program of Accountancy, School of Management, Walailak University

* Corresponding author E-mail: tapanee.t@eng.kmutnb.ac.th

Abstract

The objective of this research aimed to apply data mining techniques to investigate the market price of common stock from SET50 in the Stock Exchange of Thailand in order to forecast the market price for investment. The population consist of fifty companies. We collected the financial statements (form 56-1) all fifty companies in SET50 since 2016 to 2020. The factors or variables that we concerned are composed of seven variables: revenue, net profit, profit per shareholder's equity, shareholder's equity, number of common shares, registered capital, and dividend. We used three various data mining techniques including Generalized Regression Model, Decision Tree, and Support Vector Machine by using RapidMiner Go as a tool. The research results indicated that all seven factors or variables that we concerned significantly affect the market price. However, each data mining technique forecast different levels. We considered the values of RMSE (Root Mean Squared Error) importantly; less RMSE, less error. We found that RMSE from lowest to highest value are Decision Tree (28.19), Support Vector Machine (35.05), and Generalized Regression Model (40.78) respectively. It meant that Decision Tree is the best for forecasting in this research. Decision Tree showed the results in the form of decision tree visualization. Support Vector Machine and Generalized Regression Model showed the result in term of weight and coefficient consecutively.

Keywords : Data mining, Forecasting, Common stock price, Finance statement

1. บทนำ

ตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) ก่อตั้งขึ้นในปี พ.ศ. 2518 โดยมีวัตถุประสงค์เพื่อเป็นศูนย์กลางสำหรับการซื้อขายหลักทรัพย์สำหรับบริษัทที่ต้องการเงินทุนเพื่อขยายธุรกิจจะได้มีช่องทางระดมทุน โดยการเสนอขายหลักทรัพย์แก่ประชาชนทั่วไปโดยมีทั้งการเสนอขายหลักทรัพย์ใหม่ต่อประชาชนเป็นครั้งแรก และการเสนอขายหลักทรัพย์ทั่วไปในตลาดรองที่ให้ผู้ถือหุ้นสามารถซื้อขายหลักทรัพย์กันได้ การลงทุนในบริษัทที่จดทะเบียนในตลาดหลักทรัพย์นั้นถือเป็นการส่งเสริมการออมให้กับประชาชนทั่วไป โดยให้ผู้มีเงินออมเหลือนำมาลงทุนหาผลตอบแทนจากการถือครองหลักทรัพย์ซึ่งเสมือนเป็นผู้มีส่วนร่วมในการเป็นเจ้าของกิจการธุรกิจอุตสาหกรรมภายในประเทศ รวมถึงยังเป็นช่องทางระดมเงินทุนระยะยาวให้กับภาคธุรกิจไว้สำหรับใช้ขยายกิจการ ซึ่งถือเป็นบทบาทสำคัญที่จะส่งผลให้เกิดการพัฒนาเศรษฐกิจของประเทศต่อไป (ตลาดหลักทรัพย์แห่งประเทศไทย, 2564 : เว็บไซต์) ปัจจุบันการลงทุนในตลาดหลักทรัพย์ถือเป็นทางเลือกที่เหมาะสมสำหรับผู้ที่มีเงินออมหรือ มีความต้องการเป็นเจ้าของกิจการที่มีศักยภาพและมีแนวโน้มการเจริญเติบโตที่ดี และ ขณะนี้เนื่องจากประเทศไทยกำลังเข้าสู่สังคมผู้สูงอายุ ดังนั้นจึงควรส่งเสริมประชาชนให้ตระหนักใส่ใจและรู้จักวางแผนเงินออมและการลงทุนหุ้นสามัญในตลาดหลักทรัพย์เพื่อที่จะสามารถวางแผนการเกษียณของตนเองได้ในอนาคต (ชินกฤต วงศ์รักษ์, 2564) ซึ่ง Harlow (2011) ได้เสนอสัดส่วนการลงทุนว่าควรนำเงินไปลงทุนในหุ้นสามัญในสัดส่วนร้อยละ 5 ถึงร้อยละ 25 เพื่อใช้ยามเกษียณ อย่างไรก็ตามสิ่งที่เป็นคำถามสำคัญและนำมาสู่งานวิจัยชิ้นนี้ คือ เมื่อประชาชนสนใจที่จะลงทุนในตลาดหลักทรัพย์แล้วจะตัดสินใจได้อย่างไรถูกต้องอย่างไรว่าหุ้นตัวไหนมีความเหมาะสมในการลงทุน มีนักวิจัยพยายามหาคำตอบโดยใช้หลายกระบวนการ หลายวิธี และ หลายเทคนิค แต่ก็ยังไม่คำตอบที่สมบูรณ์ที่สุด เนื่องด้วยคุณลักษณะของตลาดหุ้นที่ข้อมูลเป็นอนุกรมเวลา (Time Series Data)

ดังนั้น จึงมีเรื่องของความไม่อยู่นิ่ง (Non-Stationary) นั่นคือในแต่ละจุดของข้อมูลมีการเปลี่ยนแปลงของค่าเฉลี่ย (Mean) ความแปรปรวน (Variance) และ ความแปรปรวนร่วมเกี่ยว (Covariance) ตลอดเวลา มีความไม่เป็นเชิงเส้น (Nonlinearity) และ มีความผันผวน (Volatility) หรือ การเคลื่อนไหวหรือการเปลี่ยนแปลงของราคาอย่างมาก (Polanco-Martinez, 2019) การพยายามหาคำตอบเรื่องนี้จึงเป็นความท้าทายอย่างมาก และในปัจจุบันด้วยความก้าวหน้าของเทคโนโลยีการประมวลผลข้อมูล การทำนายผลเริ่มมีแนวโน้มที่ดีขึ้น (Hoseinzade & Haratizadeh, 2019) ซึ่งการพยากรณ์ราคาหุ้นเป็นศาสตร์ที่ต้องอาศัยความรู้ทั้งด้านการเงินวิทยาการข้อมูลและคอมพิวเตอร์มาผสมผสานกัน

ราคาของหุ้นขึ้นอยู่กับปัจจัยใดเป็นสำคัญ และ ราคาตลาดของหุ้นควรเป็นเท่าไร มีหลายปัจจัยมากที่ควรพิจารณา ทั้งปัจจัยภายนอก เช่น เรื่องของเศรษฐกิจและการเมือง และ ปัจจัยภายในของตัวบริษัทเองด้วย เช่น งบการเงิน ตัวเลขทางบัญชี เป็นต้น ดังนั้น การวิเคราะห์ต้องพิจารณาจากข้อมูลในอดีตย้อนกลับไปแล้วจะเห็นการเคลื่อนไหวของตลาดอย่างแท้จริง (Masoud, 2017; Murkute & Tanuja, 2015) สองแนวทางหลักที่นิยมใช้ในการทำนายราคาหุ้น วิธีที่หนึ่ง คือ การวิเคราะห์เชิงเทคนิค (Technical Analysis) วิธีนี้จะใช้ข้อมูลในอดีต เช่น ราคาเปิดและราคาปิด จำนวนหรือปริมาณหุ้นที่ซื้อขาย และอื่น ๆ ในการทำนายราคาหุ้นในอนาคต ส่วนวิธีที่สอง คือ การวิเคราะห์เชิงคุณภาพ (Qualitative Analysis) จะใช้ปัจจัยภายนอกประกอบการทำนายด้วย เช่น โปรไฟล์ของบริษัท สภาพตลาด เศรษฐกิจ สังคมและการเมือง เป็นต้น และ ในปัจจุบันก็เพิ่มเติมด้วยข้อมูลที่เป็นตัวอักษร (Textual Information) จากข่าวที่เกี่ยวข้องกับการเงิน สื่อสังคมออนไลน์ต่าง ๆ และ บล็อกจากนักวิเคราะห์ข้อมูลทางเศรษฐกิจ (Hur, Manoj, & Yohanes, 2006) ปัจจุบันเมื่อข้อมูลมีปริมาณ (Volume) มากมายมากยิ่งขึ้น และมีความซับซ้อน (Variety) มากยิ่งขึ้น รวมถึงมีการเคลื่อนไหว (Velocity) อยู่เสมอ หรือ ที่นิยมเรียกว่า บิ๊กดาต้า (Big Data) หรือ ข้อมูลขนาดใหญ่ นักวิจัยจึงมองหาวิธีการช่วยแก้ปัญหาดังกล่าว โชคดีที่เทคโนโลยีของคอมพิวเตอร์ในปัจจุบันก็มีความก้าวหน้าและพัฒนามากยิ่งขึ้นเช่นกัน จึงสามารถนำวิธีการวิเคราะห์ข้อมูล (Data Analytics) หลายรูปแบบ เช่น Artificial Intelligent, Data Mining, Machine Learning หรือ Deep Learning มาช่วยในการวิเคราะห์ข้อมูลได้ สำหรับงานวิจัยนี้ได้เลือกเทคนิคการทำเหมืองข้อมูล (Data Mining) มาช่วยในการวิเคราะห์เนื่องจากวิธีนี้มีความเหมาะสมสำหรับการวิเคราะห์ข้อมูลของงบการเงินและการทำนายราคาหุ้นอย่างมากวิธีหนึ่ง

การทำเหมืองข้อมูล (Data Mining) คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ (Pattern) หรือ ความสัมพันธ์ (Relation) ที่ซ่อนอยู่ในชุดข้อมูลนั้น ๆ ซึ่งถ้าสามารถค้นพบรูปแบบหรือความสัมพันธ์นั้น ๆ หรือจะเรียกว่าความรู้ที่ได้ (Knowledge) และ นำความรู้เหล่านั้นมาใช้ให้เกิดประโยชน์ก็จะมีคุณค่ามาก เปรียบเสมือนการทำเหมืองที่ต้องขุดดิน หินทรายเป็นจำนวนมากแต่ถ้าได้ค้นพบแร่ที่สำคัญ เช่น เพชร หรือ ทอง ก็จะได้พบประโยชน์ที่มีคุณค่ามหาศาล ปัจจุบันการทำเหมืองข้อมูลถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม ตัวอย่างที่เด่นชัดของการทำเหมืองข้อมูล คือ การค้นพบรูปแบบการซื้อขายของในซูเปอร์มาร์เก็ตแห่งหนึ่ง พบว่าลูกค้าผู้ขายจะซื้อเบียร์พร้อมกับผ้าอ้อมเด็กเสมอ ๆ ในทุกเย็นวันศุกร์ จากการค้นพบนี้ทำให้ซูเปอร์มาร์เก็ตแห่งนั้นเปลี่ยนแปลงการจัดชั้นวางของ โดยนำเบียร์และผ้าอ้อมเด็กมาวางไว้ที่ชั้นวางของใกล้กัน ผลปรากฏว่าสามารถเพิ่มยอดขายได้อย่างมากมาย นี่คือนิยามของการค้นพบรูปแบบของความสัมพันธ์และนำมาประยุกต์ใช้ให้เกิดประโยชน์

ด้วยคุณประโยชน์ของการทำเหมืองข้อมูล คือ สามารถค้นหารูปแบบ (Pattern) หรือความสัมพันธ์ (Relation) ที่ซ่อนอยู่ในชุดข้อมูลที่มีขนาดใหญ่และหลากหลายหลายตัวแปรพร้อมกันทั้งข้อมูลที่มีอยู่ในอดีตและที่เป็นปัจจุบันดังนั้นงานวิจัยนี้จึง

เลือกเทคนิคการทำเหมืองข้อมูลมาทำนายราคาของหุ้นจากตัวแปรที่หลากหลายพร้อมกันโดยใช้ข้อมูลทั้งในอดีตและปัจจุบันเป็นตัวแปรทำนาย

2. การทบทวนวรรณกรรม

2.1 เอกสารงานวิจัยที่เกี่ยวข้อง

จากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องคณะผู้วิจัยพบว่า มีหลายงานวิจัยที่ทำการวิจัยเกี่ยวกับเรื่องของหุ้น ทั้งเรื่องผลการดำเนินการของบริษัทนั้น ๆ หรือ ราคาของหุ้นในกลุ่มต่าง ๆ โดยอาศัยข้อมูลทั้งปัจจัยภายใน เช่น งบการเงิน หรือ ปัจจัยภายนอก เช่น สภาพเศรษฐกิจ เป็นต้น ส่วนเครื่องมือที่นิยมใช้ได้แก่การใช้หลักการของสถิติ และ การทำเหมืองข้อมูล ตัวอย่างเช่น งานวิจัย (พิรุณพร ไพฑูรย์เจริญผล, 2563) ที่ต้องการพยากรณ์ผลการดำเนินงานของบริษัทในตลาดหลักทรัพย์ จากข้อมูลงบการเงินของบริษัทที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย ในกลุ่มอุตสาหกรรมเกษตร และกลุ่มอุตสาหกรรมอาหาร โดยใช้เทคนิคการทำเหมืองข้อมูลแบบ Deep Learning พบว่า มี 7 ตัวแปรสำคัญที่ช่วยทำนายผลการดำเนินการของกิจการได้ ได้แก่ อัตราส่วนกำไรก่อนดอกเบี้ยและภาษีต่อสินทรัพย์รวม อัตราส่วนต้นทุนกำไรจากการดำเนินงานต่อยอดขาย อัตราส่วนกำไรสุทธิต่อยอดขาย อัตราส่วนผลตอบแทนต่อผู้ถือหุ้น อัตราแลกเปลี่ยนเงินสกุลดอลลาร์สหรัฐ อัตราส่วนเงินสดต่อสินทรัพย์รวม และอัตราส่วนกำไรก่อนดอกเบี้ยและภาษีเงินได้ต่อหนี้สินรวม

งานวิจัย (ยศสยา แสงหิรัญ และ สมชาย เล็กเจริญ, 2561) ศึกษาค่าความแปรปรวนตัวแปรของหุ้นกลุ่มเทคโนโลยี พบว่า ตัวแปรที่มีผลต่อการเปลี่ยนแปลงของราคาปิดของหุ้นในแต่ละเดือนมากที่สุดคือ มูลค่าตามบัญชีต่อหุ้น: Book Value Per Share (BVPS) รองลงมาคือ ผลตอบแทนต่อสินทรัพย์: Return on Assets (ROA) และสุดท้ายคือ อัตราผลตอบแทนต่อผู้ถือหุ้น: Return On Equity (ROE) ตามลำดับ โดยเทคนิคการทำเหมืองข้อมูลที่น่าสนใจได้แก่ อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree, J48) การจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest) การแยกประเภทแบบการหาเพื่อนบ้านใกล้ที่สุด (K- Nearest Neighbor: KNN) และ โครงข่ายประสาทเทียม (Neural Network) กรณีศึกษาใช้จำนวนหุ้นทั้งหมด 36 หุ้น ระยะเวลาตั้งแต่ปี พ.ศ. 2556 ถึง 2560 พบว่า การแยกประเภทแบบการหาเพื่อนบ้านใกล้ที่สุด (K- Nearest Neighbor: KNN) มีประสิทธิภาพดีที่สุด

งานวิจัยของ (อริพัทธ์ วจนวุฒิจิตติคุณ, 2554) เป็นเรื่องที่ใกล้เคียงกัน คือ การหาปัจจัยที่มีผลกระทบต่อราคาหลักทรัพย์ว่ามีปัจจัยใดบ้างแต่งานวิจัยนี้เน้นในกลุ่มธนาคารพาณิชย์ในตลาดหลักทรัพย์ โดยศึกษาปัจจัยจำนวน 4 ตัวแปร ได้แก่ อัตราเงินเฟ้อ อัตราดอกเบี้ยเงินฝากที่แท้จริง เงินปันผลเฉลี่ย และ ผลิตภัณฑ์มวลรวมประชาชาติที่แท้จริง ใช้ข้อมูลจำนวน 15 ไตรมาส โดยใช้วิธีการวิเคราะห์แบบการถดถอยพหุเชิงเส้น (Multiple Linear Regression) ด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square) พบว่า มีเพียงปัจจัยเดียวเท่านั้นที่มีผลต่อผลกระทบต่อราคาหลักทรัพย์ในกลุ่มธนาคารพาณิชย์ คือ ผลิตภัณฑ์มวลรวมประชาชาติที่แท้จริง

นอกจากนี้ มีงานวิจัยที่ศึกษาปัจจัยทางเศรษฐกิจที่มีผลต่อราคาหุ้นในแต่ละเดือนของ (ณัฐรฐา สันทัดงาน และ กิตติพันธ์ คงสวัสดิ์เกียรติ, 2559) แต่วิจัยกับบริษัทเดียวคือ บริษัท เด็มโก้ (DEMCO) จำกัด (มหาชน) ปัจจัยที่น่าสนใจพิจารณา ได้แก่ อัตราเงินเฟ้อ (IR) อัตราดอกเบี้ยเงินฝากประจำ 3 เดือน (DR: FIX3) อัตราแลกเปลี่ยนเงินบาทต่อดอลลาร์สหรัฐอเมริกา (EX) ดัชนีการลงทุนภาคเอกชน (PII) และดัชนีค่าเงินบาทที่แท้จริง (REER) ข้อมูลตั้งแต่ปี พ.ศ. 2553 ถึง 2557 (60 เดือน) โดยใช้วิธีการวิเคราะห์แบบการถดถอยพหุคูณ (Multiple Regression) ด้วยวิธีแบบขั้นต่อน (Stepwise

Regression) พบว่าตัวแปรที่อธิบายราคาหุ้นได้ดี คือ อัตราดอกเบี้ยเงินฝากประจำ 3 เดือน (DR:FIX3) อัตราแลกเปลี่ยนบาทต่อดอลลาร์สหรัฐ (EX) และ ดัชนีค่าเงินบาทที่แท้จริง (REER) โดยทั้งสามตัวแปรมีความสัมพันธ์ในทิศทางเดียวกัน

มีงานวิจัยที่พยากรณ์แนวโน้มราคาหลักทรัพย์ (ปรเมษฐ์ ธันวานนท์ และคณะ, 2560) โดยศึกษาปัจจัย 4 ปัจจัยที่เกี่ยวข้องใกล้เคียงกับหลักทรัพย์นั้น ๆ คือ ข้อมูลราคาซื้อขายหลักทรัพย์ ข้อมูลดัชนีชี้วัดทางเทคนิค(Indicators Data) จำนวนวันที่ถือ (Holding Days) และ จำนวนวันที่ใช้ในการคำนวณค่าดัชนีชี้วัดทางเทคนิค (Indicator Days) โดยใช้เทคนิคการวิเคราะห์การเรียนรู้ของเครื่อง (Machine Learning) จำนวน 4 เทคนิค ได้แก่ Random Forest, Support Vector Machine, K-Nearest Neighbors และ Naïve Bayes ร่วมกับเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Model) โดยใช้ข้อมูลของตลาดหลักทรัพย์แห่งประเทศไทยตั้งแต่ปี พ.ศ. 2554 ถึง 2559 ผลที่ได้พบว่า เทคนิค Ensemble Model สามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ถึง 5%-14%

2.2 ทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล (Data Mining) มีวัตถุประสงค์หลัก คือ ค้นหารูปแบบ (Pattern) หรือความสัมพันธ์ (Relation) ที่ซ่อนอยู่ และ นำสิ่งที่ค้นพบนั้นไปใช้ประโยชน์ ซึ่งเทคนิคในการทำเหมืองข้อมูลสามารถจัดได้เป็น 3 ประเภทหลัก ดังนี้ ได้แก่

1. การหาความสัมพันธ์ของข้อมูล (Association Rule) เป็นการค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปด้วยกัน ส่วนใหญ่ใช้ในงานซื้อขายสินค้า หรือ เรียกว่า Basket Analysis เช่น พบว่ามีการซื้อเบียร์พร้อมผ้าอ้อมด้วยกัน กฎจะทำการวัดโดยได้ค่าแรก คือ ค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ หรือเป็นเปอร์เซ็นต์ที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่ คือ ค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่ถูกถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ ซึ่งขั้นตอนในการหาหลายวิธีด้วยกัน แต่วิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลาย คือ อัลกอริทึมอะพริออริ (Apriori Algorithm) และ เอฟพี โกรท (FP-Growth Algorithm) (สุภาพรณ คงฉวีพรรณ, 2562)

2. การจัดกลุ่ม (Clustering) คือ การจัดกลุ่มข้อมูลเพื่อค้นหากลุ่มที่สามารถยอมรับได้เพื่อจัดเข้ากลุ่ม กล่าวคือ กลุ่มของวัตถุมีการสร้างขึ้นโดยเปรียบเทียบวัตถุที่มีความเหมือนกันจัดเข้ากลุ่มเดียวกัน เช่น การจัดกลุ่มของลูกค้าที่มีความหลากหลาย เช่น อายุ อาชีพ เพศ รายได้ สถานภาพ งานอดิเรก จังหวัดที่อยู่ ความชอบ ไลฟ์สไตล์ ฯลฯ ให้เป็นกลุ่มต่าง ๆ เพื่อสะดวกในการจัดโปรโมชันต่าง ๆ ซึ่งมีหลากหลายวิธีด้วยกันในการจัดกลุ่ม เช่น อัลกอริทึมเค-มีนส์ (K-means Algorithm) การวิเคราะห์แบบลำดับขั้น (Hierarchical Analysis) และ ดีบีSCAN (DBSCAN) (สุทธิพงศ์ ผ่องแผ้ว, 2561)

3. การแบ่งประเภท/การจำแนกข้อมูล (Classification) เป็นการจัดแบ่งประเภทของข้อมูล โดยหาชุดต้นแบบเพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบ เช่น ต้นไม้ตัดสินใจ (Decision Trees) หรือ โครงข่ายประสาท (Neural Networks) เป็นต้น

สำหรับวัตถุประสงค์ของงานวิจัยนี้ คือ ต้องการพยากรณ์ (Prediction) ราคาหุ้นว่าควรจะเป็นเท่าไรเพื่อตัดสินใจในการจะลงทุนและค้นหาว่าตัวแปรใดที่เกี่ยวข้องกับราคาของหุ้น ดังนั้น จึงเลือกการทำเหมืองข้อมูลแบบการแบ่งประเภท (Classification) มาประยุกต์ใช้ โดยการพยากรณ์ (Prediction) สามารถแบ่งออกเป็น 2 รูปแบบหลัก ๆ คือ

1. การพยากรณ์ด้วยการวิเคราะห์การถดถอย (Regression-Based Prediction) เป็นการหาความสัมพันธ์ระหว่างตัวแปร หรือ คุณลักษณะของข้อมูล ผลของการวิเคราะห์จะเป็นข้อมูลเชิงปริมาณ (Quantitative Data) การวิเคราะห์

การถดถอยมีหลายรูปแบบ เช่น การวิเคราะห์การถดถอยเชิงเส้นเชิงเดียว (Simple Linear Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระเพียงหนึ่งตัว (x) กับตัวแปรตาม (y) โดยความสัมพันธ์คือ $y = a + bx + e$ ส่วนการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัว ($x_1, x_2, x_3, \dots$) กับตัวแปรตาม (y) โดยความสัมพันธ์คือ $y = a + bx_1 + cx_2 + dx_3 + \dots + e$ เป็นต้น

2. การพยากรณ์ด้วยการจำแนกข้อมูล (Classification Based Prediction) เป็นกระบวนการสร้างโมเดลการเรียนรู้จากข้อมูลที่มีอยู่โดยการสร้างเป็นกฎหรือเงื่อนไขเพื่อช่วยในการตัดสินใจ ข้อมูลของการวิเคราะห์จะเป็นข้อมูลเชิงคุณภาพ (Qualitative Data) ซึ่งการพยากรณ์ด้วยจำแนกข้อมูลก็ยังมีหลากหลายเทคนิค เช่น

2.1 ต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนี้จะให้ผลลัพธ์ในลักษณะโครงสร้างของต้นไม้ ภายในต้นไม้ประกอบไปด้วยโหนด (Node) กิ่งของต้นไม้ (Branch) และ ใบ (Leaf) แต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัวทดสอบ กิ่งของต้นไม้แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ ส่วนใบ (Leaf) อยู่ด้านล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (Class) นั่นคือผลลัพธ์จากการพยากรณ์ ข้อดีของวิธีนี้คือการแสดงผล (Visualization) ให้เข้าใจได้ดี และมีความแม่นยำ

2.2 การสุ่มป่าไม้ (Random Forest) แนวคิดของเทคนิคนี้ คือ การสร้างโมเดลด้วยวิธีการ Decision Tree ขึ้นมาหลาย ๆ โมเดลโดยวิธีการสุ่มตัวแปรแล้วนำผลที่ได้แต่ละโมเดลมารวมกัน พร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย

2.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM) เป็นการหาขอบเขตการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยพยายามสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงระหว่างกึ่งกลางของกลุ่มข้อมูลให้มีระยะห่างระหว่างขอบเขตของทั้งสองกลุ่มมากที่สุด โดยปกติ SVM ถูกนำมาใช้กับข้อมูลเชิงเส้น แต่ในความเป็นจริงข้อมูลมักไม่เป็นเชิงเส้นจึงแก้ปัญหาโดยนำ Kernel Function มาใช้เพื่อสามารถเปลี่ยนข้อมูลที่มีมิติต่ำกว่าให้มีมิติสูงขึ้นได้ เช่น การแบ่งข้อมูลแบบ Linear, Polynomial, Radial Basis Function เป็นต้น

2.4 เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) เป็นวิธีการจำแนกกลุ่มของข้อมูล การสกัดข้อมูลจำแนก เหมาะสำหรับการจำแนกกลุ่มของข้อมูลที่มีลักษณะกระจาย โดยจะหาระยะห่างของข้อมูลคุณลักษณะที่ต้องการพยากรณ์กับข้อมูลผลลัพธ์ที่มีอยู่ก่อนหน้า โดยจะเลือกข้อมูลที่อยู่ใกล้ที่สุดจำนวน K ตัว จากนั้นสกัดข้อมูลจำแนกและพยากรณ์ผล วิธีการนี้มีประสิทธิภาพดี ถ้า Training Data มีขนาดใหญ่

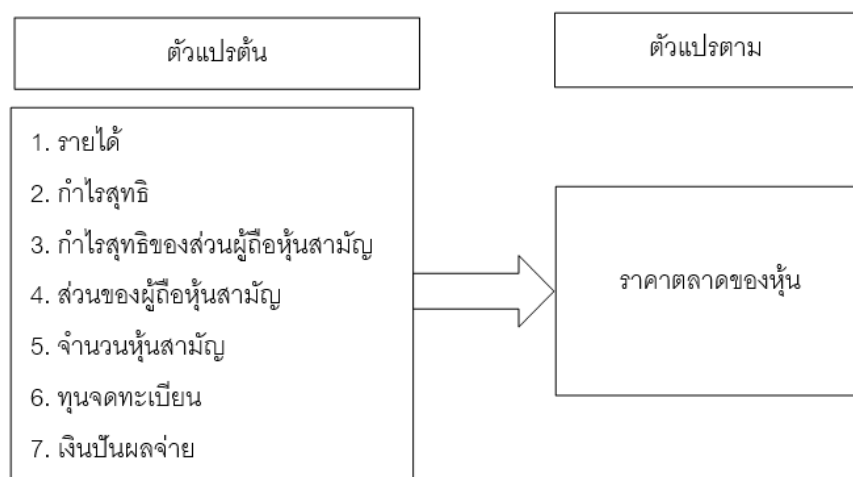
2.5 การเรียนรู้เชิงลึก (Deep Learning) เป็นวิธีการการเรียนรู้โดยอาศัยพื้นฐานของโครงข่ายประสาทเทียมและการเรียนรู้เชิงคุณลักษณะโดยคำว่าลึก คือ การมีชั้นของโครงข่ายหลายชั้น ทำให้มีประสิทธิภาพมากขึ้นและเข้าใจโครงสร้างที่ชัดเจนขึ้น

2.6 นาอ็ฟ เบย์ (Naïve Bayes) เป็นการจำแนกประเภทข้อมูล และสมมติฐานที่ให้การเกิดของเหตุการณ์ต่าง ๆ โดยใช้หลักความน่าจะเป็นตามแนวคิดของ Bayes' Theorem ข้อดีของวิธีการนี้ คือ เข้าใจและนำไปใช้งานได้ง่าย

3. สมมติฐานของการวิจัย

คณะผู้วิจัยต้องการศึกษาว่าตัวแปรต้นจำนวน 7 ตัวแปรที่ได้มาจากการเงินของบริษัทนั้น ๆ ได้แก่ รายได้ กำไรสุทธิ กำไรสุทธิของส่วนผู้ถือหุ้นสามัญ ส่วนของผู้ถือหุ้นสามัญ จำนวนหุ้นสามัญ ทุนจดทะเบียน และ เงินปันผลจ่าย จะสามารถทำนายหรือมีผลต่อราคาตลาดของหุ้นหรือไม่อย่างไร

ภาพประกอบ 1 กรอบแนวคิด



ราคาตลาดของหุ้น (Market Price) หมายถึง ราคาที่ตกลงซื้อขายหุ้นกันตามความต้องการ หากราคาในการเสนอซื้อ มีมากกว่าราคาเสนอขาย จะทำให้ผลตอบแทนต่อผู้ถือหุ้นมีการตอบแทนที่สูงขึ้น แต่ถ้าความต้องการเสนอขายมีมากกว่าความต้องการเสนอซื้อจะส่งผลให้ราคาตลาดของหุ้นนั้นต่ำลง

4. วิธีการดำเนินงานวิจัย

1. ประชากรในงานวิจัยนี้ คือ บริษัทที่จดทะเบียนในตลาดหลักทรัพย์ (SET) และ กลุ่มตัวอย่าง คือ กลุ่มบริษัท ใน SET 50 โดย SET50 คือ หุ้น 50 อันดับแรกที่ผ่านมาเกณฑ์ที่กำหนดของตลาดหุ้น เช่น ADVANC, AOT, AWC, BAM, BBL, BDMS, BGRIM, BTS, CPALL, CPF, CPN, DELTA, DTAC, EGCO, GULF, HMPRO, INTUCH, KBANK, KTB, LH, OR, OSP, PTT, SCB, SCC, TISCO, TMB, TOA, TOP, TRUE เป็นต้น

2. เครื่องมือที่ใช้ในการศึกษา

ข้อมูลดิบ: งานวิจัยนี้ใช้ข้อมูลรายงานประจำปีโดยอ้างอิงจากงบการเงินในรูปแบบฟอร์ม 56-1 โดยแบบฟอร์ม 56-1 คือรายงานประจำปีของบริษัทที่จดทะเบียนที่อยู่ในตลาดหลักทรัพย์แห่งประเทศไทย ซึ่งบริษัทจะต้องจัดทำเพื่อเผยแพร่บน เว็บไซต์ของตลาดหลักทรัพย์ทุกปีซึ่งจะแสดงผลในช่วงไตรมาสแรกของปีถัดไป (ส่วนมากจะออกประมาณช่วงเดือนมีนาคมถึง เดือนเมษายน) โดยเก็บข้อมูลตั้งแต่ปี พ.ศ. 2559 ถึง 2563 โดยตัวแปรตามที่ต้องการทำนาย คือ ราคาตลาดของหุ้นสามัญที่เกิดขึ้น ส่วนตัวแปรอิสระหรือตัวแปรต้นที่ศึกษา ได้แก่ รายได้ กำไรสุทธิ กำไรต่อส่วนของผู้ถือหุ้นสามัญ ส่วนของผู้ถือหุ้นสามัญ จำนวนหุ้นสามัญ ทุนจดทะเบียน และการจ่ายเงินปันผล ส่วนการวิเคราะห์ข้อมูลครั้งนี้คณะผู้วิจัยเลือกใช้โปรแกรม RapidMiner Go

โปรแกรม RapidMiner เป็นซอฟต์แวร์ด้านวิทยาการคอมพิวเตอร์ (Data Science) ใช้สำหรับการเตรียมข้อมูล (Data Preparation) การเรียนรู้ของเครื่อง (Machine Learning) การเรียนรู้เชิงลึก (Deep Learning) การทำเหมืองข้อมูล (Data Mining) และการวิเคราะห์การทำนาย (Predictive Analysis) คุณลักษณะเด่น คือ สามารถจัดการข้อมูล และลดข้อผิดพลาด ในการจัดการรวมไปถึงการโหลดและการแปลงข้อมูล (ETL) การประมวลผลล่วงหน้า การวาดภาพจากข้อมูล การวิเคราะห์เชิง

พยากรณ์ การสร้างแบบจำลองทางสถิติ การประเมินผล และการปรับใช้ ส่วน RapidMiner Go เป็นส่วนหนึ่งของโปรแกรม RapidMiner ที่มีส่วนติดต่อผู้ใช้ (User Interface, UI) ที่ช่วยจัดการและวิเคราะห์ข้อมูลให้สะดวกมากยิ่งขึ้น

3. ขั้นตอนการดำเนินงานวิจัย คณะผู้วิจัยได้นำข้อมูลงบการเงินของบริษัทที่จดทะเบียนที่อยู่ในตลาดหลักทรัพย์แห่งประเทศไทยที่เป็นหุ้น SET50 มาทำนายราคาตลาด โดยเลือกใช้ตัวแปรที่คิดว่าจะมีผลต่อราคาตลาดซึ่งจากการทบทวนวรรณกรรมพบว่าตัวแปรที่เลือกมาเหล่านี้มีแนวโน้มที่จะส่งผลต่อราคาตลาดแต่ยังไม่มีการวิเคราะห์ในภาพรวมร่วมกัน โดยตัวแปรเหล่านี้ ได้แก่ รายได้ กำไรสุทธิ กำไรต่อส่วนของถือหุ้นสามัญ ส่วนของถือหุ้นสามัญ จำนวนหุ้นสามัญทุนจดทะเบียน และการจ่ายเงินปันผล ตั้งแต่ปี พ.ศ. 2559 ถึง 2563 มาใช้ในการวิเคราะห์ ซึ่งขั้นตอนการดำเนินการเริ่มจากคณะผู้วิจัยจัดเก็บและรวบรวมข้อมูลดิบจำนวนห้าปี จากนั้นได้นำข้อมูลดิบมาทำความสะอาด (Data Cleansing) ปรับแต่งจากนั้นจึงนำมาวิเคราะห์ด้วยโปรแกรม RapidMiner Go ต่อไป

5. ผลลัพธ์การวิจัย

เมื่อวิเคราะห์ข้อมูลจำนวน 50 บริษัท โดยใช้ตัวแปรต้น 7 ตัว ตั้งแต่ปี พ.ศ. 2559 ถึง 2563 ด้วยโปรแกรม RapidMiner Go ซึ่งเป็นหนึ่งในโปรแกรมสำหรับการทำเหมืองข้อมูลที่มีชื่อเสียงและได้รับความนิยมอย่างมากในปัจจุบัน ที่สำคัญมีเวอร์ชันที่เป็นโอเพ่นซอร์ส (Open Source) ผลลัพธ์แรกที่ได้รับ คือ ค่า RMSE (Root Mean Squared Error) ของแต่ละเทคนิคของการทำเหมืองข้อมูล RMSE คือ การทำ Square Root (SQRT) ค่า MSE (Mean Squared Error) เพื่อให้ได้ค่า Loss ที่มีหน่วยเดียวกับตัวแปรตาม เหตุผลที่ต้องทำ SQRT เพราะว่าการยกกำลังสองค่า Error ก่อน แล้วจึงหาค่าเฉลี่ยจะทำให้หน่วยเปลี่ยนไปจากเดิมตามสมการด้านล่าง ค่า RMSE จะทำให้การแปลผลลัพธ์สะดวกและชัดเจนมากยิ่งขึ้น

$$RMSE = \sqrt{\frac{1}{n} * \sum (prediction - actual)^2}$$

ตาราง 1 แสดงค่า RMSE (Root Mean Squared Error) ของการวิเคราะห์ข้อมูลด้วยเทคนิค Decision Tree, Support Vector Machine และ Generalized Linear Regression Model

ตาราง 1 ค่า RMSE (Root Mean Squared Error) ของการวิเคราะห์ข้อมูล

เทคนิคการทำเหมืองข้อมูล (Data Mining) ที่เลือกใช้	ค่า RMSE (Root Mean Squared Error)
ต้นไม้ตัดสินใจ (Decision Tree หรือ DT)	28.19
ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine หรือ SVM)	35.05
การวิเคราะห์การถดถอยเชิงเส้น (Generalized Linear Regression Model)	40.78

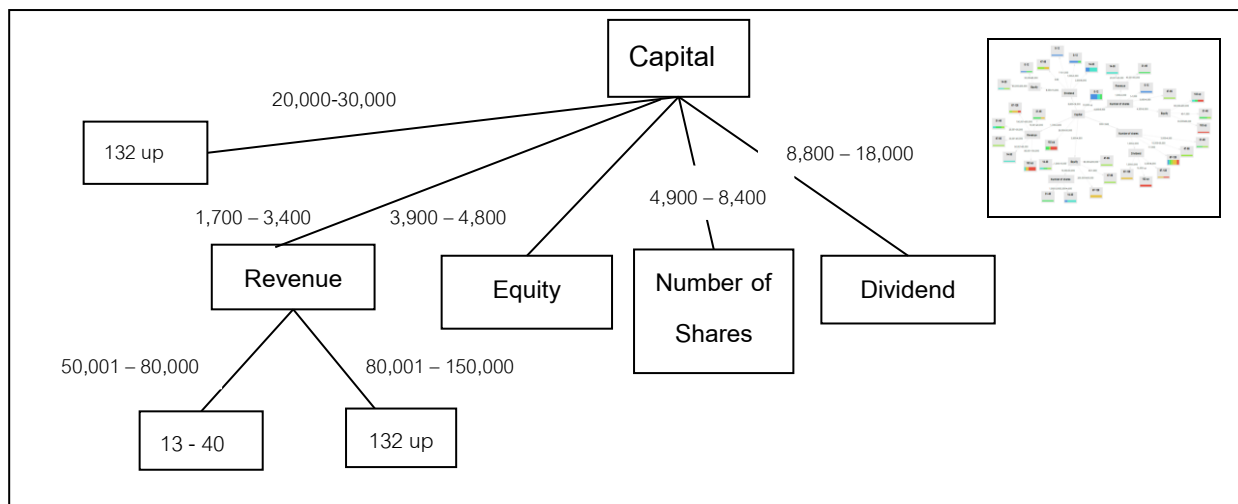
ผลที่ได้พบว่า ค่า RMSE ของ Decision Tree มีค่าเท่ากับ 28.19 แปลความได้ว่า โดยเฉลี่ยโมเดลสามารถทำนายตัวแปรตาม (y) ผิดไป +/- 28.19 จุด โดย RMSE ยังมีค่าต่ำยิ่งดี ถ้าเท่ากับ 0 แปลว่าโมเดลทำนายค่า y ได้ถูกต้อง 100% ซึ่งในทางปฏิบัติโอกาสที่จะหาค่าโมเดลให้ Loss = 0 เป็นไปได้อย่างยากมาก ส่วน Support Vector Machine มีค่าเท่ากับ 35.05 และ Generalized Linear Regression Model มีค่าเท่ากับ 40.78 ตามลำดับ

5.1 ผลลัพธ์ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจมีการวัดค่าความสามารถของการจัดกลุ่มของแต่ละคุณลักษณะ ตัวแปร หรือปัจจัย ได้แก่ ค่าดัชนีจีนิ (Gini Index) และ ค่าเอนโทรปี (Entropy) (ภรณ์ยา ปาลวิสุทธิ, 2559) ค่าดัชนีจีนิเป็นค่าที่บ่งบอกว่าคุณลักษณะหรือปัจจัยใดควรนำมาใช้เป็นคุณลักษณะในการแบ่งกลุ่ม ตัวอย่างอัลกอริทึมที่ใช้ เช่น J48 และ CART ส่วนค่าเอนโทรปีเป็นการวัดความไม่แน่นอน (Randomness) ของข้อมูล ซึ่งในการทำ Model Classification นั้นสิ่งที่ต้องการ คือ การทำนายคลาสของ Target Variable ให้แม่นยำมากที่สุด นั่นคือ ต้องการลดความไม่แน่นอนให้ได้มากที่สุด หรือพยายามแยกคลาสของ Target Variable เพื่อให้ได้สัดส่วนของคลาสใดคลาสหนึ่งมากที่สุด เมื่อคณะผู้วิจัยนำข้อมูลดิบมาทำความสะอาด (Data Cleansing) แล้วและเนื่องจากข้อมูลในงานวิจัยนี้เป็นข้อมูลเชิงปริมาณ คณะผู้วิจัยจึงนำข้อมูลเชิงปริมาณเหล่านี้มาจัดกลุ่ม (Categorize) ใหม่เพื่อให้ข้อมูลอยู่ในช่วงที่เหมาะสม เพื่อให้สามารถที่จะทำนายและแสดงผลได้อย่างเหมาะสม

ผลลัพธ์ด้วยวิธีต้นไม้ตัดสินใจ พบว่า รากของต้นไม้ตัดสินใจในงานวิจัยนี้ คือ ทุนจดทะเบียน (Capital) ส่วนโหนด (Node) ในการตัดสินใจตัวถัดไป ได้แก่ รายได้ (Revenue) จำนวนหุ้นสามัญ (Equity) เงินปันผล (Dividend) และ ส่วนของผู้ถือหุ้นสามัญ (Number of shares) ตามภาพประกอบ 1 (ส่วนรูปภาพขนาดเล็กด้านขวามือของภาพเป็นผลลัพธ์ที่ได้จากโปรแกรม Rapid Miner ซึ่งผลลัพธ์มีความละเอียดมากเมื่อแสดงด้วยรูปภาพ ฯ เดียวจะไม่ค่อยคมชัด) ดังนั้น คณะผู้วิจัยจึงขออธิบายตามภาพประกอบ 1 ซึ่งเป็นหลักจะเห็นผลลัพธ์ที่เด่นชัดกว่า

ภาพประกอบ 2 ผลลัพธ์จากวิธีต้นไม้ตัดสินใจ (Decision Tree)

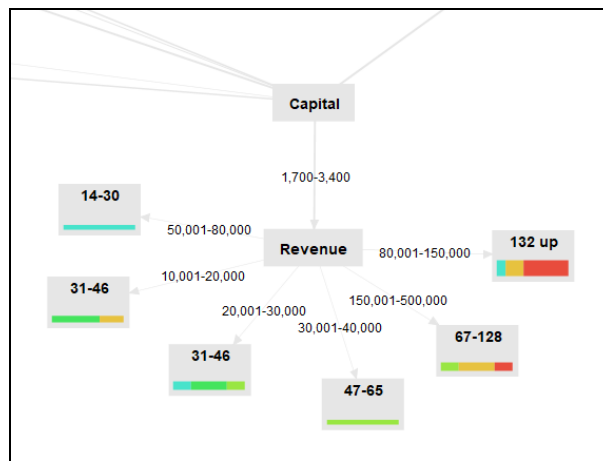


การพิจารณาต้นไม้ตัดสินใจในภาพประกอบ 2 จะเริ่มจากรากของต้นไม้ หรือ ตัวแปรที่มีผลต่อราคาตลาดของหุ้นที่สำคัญเป็นลำดับแรกซึ่งในที่นี้คือ ทุนจดทะเบียน (Capital) ผลที่ได้พบว่า ถ้าทุนจดทะเบียนอยู่ในช่วง 20,000 – 30,000 ล้านบาท ราคาตลาดของหุ้นจะอยู่ที่ 132 บาทขึ้นไป ส่วนถ้าทุนจดทะเบียนที่มีทุนอยู่ในช่วงระหว่าง 1,700 – 3,400 ล้านบาท จะต้องพิจารณาตัวแปรรายได้ (Revenue) เป็นลำดับที่สองต่อไป ซึ่งถ้ารายได้ของบริษัทอยู่ที่ 50,001-80,000 ล้านบาท ราคาหุ้นจะอยู่ที่ 14-30 บาท แต่ถ้ารายได้อยู่ที่ 80,001- 150,000 ล้านบาท ราคาจะอยู่ที่ 132 บาทขึ้นไป สำหรับบริษัทที่มีทุนจดทะเบียนที่มีทุนอยู่ในช่วงระหว่าง 3,900 – 4,800 ล้านบาทนั้น ต้องพิจารณาตัวแปรจำนวนหุ้น (Equity) เป็นลำดับถัดไป และถ้าบริษัทที่มีทุนจดทะเบียนอยู่ในช่วงระหว่าง 4,900 – 8,400 ล้านบาทนั้น ให้พิจารณาสัดส่วนของผู้ถือหุ้นสามัญ (Number of Shares) เป็นลำดับ

ถัดไปแทน แต่ถ้าบริษัทมีทุนจดทะเบียน 8,800 – 18,000 ล้านบาทนั้น จะต้องพิจารณาเงินปันผล (Dividend) เป็นลำดับถัดไป แทนเช่นกัน

ตัวอย่างการพิจารณาต้นไม้ตัดสินใจโดยลงรายละเอียดแสดงในรูปภาพประกอบ 3 การพิจารณาจะเริ่มจากรากของต้นไม้ คือ ทุนจดทะเบียน (Capital) เป็นลำดับแรก เมื่อทุนจดทะเบียนอยู่ระหว่าง 1,700 – 3,400 ล้านบาท จะต้องพิจารณาตัวแปรรายได้ (Revenue) เป็นลำดับที่สองซึ่งถ้ารายได้ของบริษัทอยู่ที่ 50,001-80,000 ล้านบาท ราคาหุ้นจะอยู่ที่ 14-30 บาท และถ้ารายได้อยู่ที่ 10,001- 30,000 ล้านบาท ราคาจะอยู่ที่ 31 - 46 บาท ถ้ารายได้อยู่ที่ 30,001- 40,000 ล้านบาท ราคาจะอยู่ที่ 47 - 65 บาท ส่วนถ้ารายได้อยู่ที่ 80,001- 150,000 ล้านบาท ราคาของหุ้นจะมากกว่า 132 บาทขึ้นไป เป็นต้น

ภาพประกอบ 3 ตัวอย่างผลลัพธ์จากวิธีต้นไม้ตัดสินใจ (Decision Tree) ในรายละเอียด



5.2 ผลลัพธ์จากวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM)

ผลลัพธ์จากวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM) จะแสดงผลเป็นค่าน้ำหนัก (Weight, W) ของตัวแปรต้นที่มีผลต่อตัวแปรตาม ยิ่งค่าน้ำหนักของตัวแปรใดมีค่ามากเท่าไรยิ่งมีผลกระทบต่อตัวแปรตามมากเท่านั้น ผลจากงานวิจัยนี้พบว่า ตัวแปรที่มีค่าน้ำหนักมากที่สุดที่มีผลต่อราคาตลาดของหุ้น คือ จำนวนหุ้น แต่ค่าที่ได้มีค่าติดลบ (-217.86) แสดงว่ายิ่งมีจำนวนหุ้นยิ่งมากเท่าไรราคาตลาดของหุ้นจะยิ่งลดต่ำลงเท่านั้น ส่วนค่าน้ำหนักตัวแปรที่มีค่าสูงถัดมา คือ เงินปันผลและมีค่าเป็นบวก (122.42) แสดงว่ายิ่งมีการจ่ายเงินปันผลมากเท่าไรยิ่งส่งผลให้ราคาตลาดของหุ้นสูงตามไปด้วย ส่วนค่าน้ำหนักตัวแปรตัวถัดมาคือทุนจดทะเบียน (96.45) กำไรสุทธิ (16.38) กำไรต่อส่วนผู้ถือหุ้นสามัญ(9.40) และ รายได้ (5.56) ซึ่งมีค่าเป็นบวก นั่นแสดงว่าบริษัทยิ่งทำกำไร และรายได้ มากเท่าไรยิ่งส่งผลให้ราคาตลาดของหุ้นสูงตามไปด้วยนั่นเอง ส่วนตัวแปรสุดท้าย คือ ส่วนผู้ถือหุ้นสามัญ (-1.28) และมีค่าเป็นลบ หมายความว่ายิ่งมีส่วนของผู้ถือหุ้นเป็นจำนวนมากราคายิ่งต่ำลง แต่ในกรณีนี้จะพบว่าค่าน้ำหนักของส่วนผู้ถือหุ้นมีค่าน้อยมากถ้าเทียบกับค่าน้ำหนักของจำนวนหุ้นดังที่อธิบายข้างต้น ค่าน้ำหนักของตัวแปรต้นแสดงในตาราง 2 ดังต่อไปนี้

ตาราง 2 ค่าน้ำหนักของตัวแปรต้น

ตัวแปรต้น	ค่าน้ำหนัก (Weight)
จำนวนหุ้นสามัญ	-217.86
เงินปันผลจ่าย	122.42
ทุนจดทะเบียน	96.45
กำไรสุทธิ	16.38
กำไรสุทธิต่อส่วนผู้ถือหุ้นสามัญ	9.40
รายได้	5.56
ส่วนผู้ถือหุ้นสามัญ	-1.28

5.3 ผลลัพธ์จากวิธี Generalized Multiple Linear Regression Model

วิธีนี้เป็นการหาความสัมพันธ์ระหว่างตัวแปรต้น ($x_1, x_2, x_3, \dots$) และตัวแปรตาม (y) ผลของการวิเคราะห์จะเป็นข้อมูลเชิงปริมาณ (Quantitative Data) ซึ่งความสัมพันธ์จะอยู่ในรูปสมการ คือ $y = a + bx_1 + cx_2 + dx_3 + \dots + e$ โดยค่า a คือ จุดตัดแกน (Intercept) ส่วน $b, c, d, \dots$ คือ ค่าสัมประสิทธิ์ (Coefficient) ซึ่งในกรณีที่ตัวแปรต้นมีหน่วยที่ต่างกันจะเลือกใช้ค่าสัมประสิทธิ์มาตรฐาน (Standard Coefficient) เพราะค่าสัมประสิทธิ์มาตรฐาน คือ ค่าที่คำนวณออกมาเพื่อปรับและแก้ไขข้อจำกัดในการเปรียบเทียบของตัวแปรที่หน่วยต่างกัน เช่น จำนวนหุ้น และ จำนวนบาท เป็นต้น ค่าสัมประสิทธิ์ของตัวแปรต้นที่ได้รับสามารถนำมาเขียนสมการ และ แสดงในตารางที่ 3 ดังนี้

ราคาตลาดของหุ้น = $74.76 - 217.86 \times (\text{ส่วนของผู้ถือหุ้น}) + 122.42 \times (\text{รายได้}) + 96.45 \times (\text{กำไรสุทธิ}) - 16.38 \times (\text{ทุนจดทะเบียน}) - 9.40 \times (\text{จำนวนหุ้น}) + 5.56 \times (\text{เงินปันผลจ่าย}) - 1.28 \times (\text{กำไรสุทธิต่อส่วนผู้ถือหุ้น})$

ตาราง 3 ค่าสัมประสิทธิ์มาตรฐานของตัวแปรต้น

ตัวแปรต้น	ค่าสัมประสิทธิ์มาตรฐาน
ส่วนของผู้ถือหุ้นสามัญ	-217.86 *
รายได้	122.42**
กำไรสุทธิ	96.45***
ทุนจดทะเบียน	-16.38****
จำนวนหุ้น	-9.40*****
เงินปันผลจ่าย	5.56
กำไรสุทธิต่อส่วนผู้ถือหุ้นสามัญ	-1.28

จากค่าสัมประสิทธิ์สามารถแปลความได้ว่า ตัวแปรส่วนของผู้ถือหุ้นมีผลต่อราคาตลาดมากที่สุดเพราะค่าสัมประสิทธิ์มาตรฐานมีค่ามากที่สุด แต่เป็นไปในทางตรงข้ามเพราะมีเครื่องหมายติดลบ (-217.86) แสดงว่ายิ่งมีส่วนของผู้ถือหุ้นยิ่งมาก ราคาตลาดก็จะลดลง เช่นเดียวกับทุนจดทะเบียน (-16.38) จำนวนหุ้นสามัญ (-9.4) และ กำไรสุทธิต่อส่วนผู้ถือหุ้น (-1.28) ยิ่งติดลบมากเท่าไรราคาตลาดของหุ้นยิ่งลดต่ำลง แต่เป็นที่น่าสนใจแม้ว่าสัมประสิทธิ์ของกำไรสุทธิต่อส่วนผู้ถือหุ้นจะติดลบ

แต่ก็มีค่าน้อยมาก (-1.28) เมื่อเทียบกับส่วนของผู้ถือหุ้น (-217.86) อย่างไรก็ตาม ในทางตรงกันข้ามกับตัวแปรในส่วนของการรายได้ (122.42) กำไรสุทธิ (96.45) และ เงินปันผลจ่าย (5.56) ตัวแปรเหล่านี้ยิ่งมีค่ามากยิ่งส่งผลให้ราคาตลาดของหุ้นมีค่ามากตามไปด้วย

6. สรุปและอภิปรายผล

การศึกษาวิจัยครั้งนี้ต้องการหาปัจจัยที่ส่งผลต่อราคาตลาดของหุ้นสามัญของบริษัทในกลุ่ม SET50 ในตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้เทคนิคการทำเหมืองข้อมูลในการพยากรณ์ (Prediction) โดยใช้ 3 เทคนิคหลักดังนี้ คือ การพยากรณ์ด้วย Generalized Multiple Linear Regression Model, Decision Tree, และ Support Vector Machine ผลที่ได้พบว่าค่า RMSE ของเทคนิค Decision Tree มีค่าเท่ากับ 28.19 Support Vector Machine เท่ากับ 35.05 และ Generalized Regression Model เท่ากับ 40.78 ตามลำดับ ดังนั้น Decision Tree จึงให้ผลการทำนายที่มีความถูกต้องมากที่สุด ตัวแปรสำคัญที่มีผลกระทบต่อราคาตลาดของหุ้นสามัญ คือ ตัวแปรทั้ง 7 ตัวที่คณะผู้วิจัยนำมาศึกษา แต่ความรุนแรงของผลกระทบมีค่าที่แตกต่างกัน บางตัวแปรส่งผลกระทบมาก บางตัวแปรส่งผลกระทบน้อย

สรุปผลจากการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ พบว่า ราก (Root) ของต้นไม้ตัดสินใจในงานวิจัยนี้ คือ ทุนจดทะเบียน (Capital) ทุนจดทะเบียนเป็นปัจจัยแรกที่ต้องพิจารณาเมื่อต้องการพยากรณ์ราคาตลาดของหุ้น ส่วนโหนด (Node) ในการตัดสินใจตัวถัดไป ได้แก่ รายได้ (Revenue) จำนวนหุ้นสามัญ (Equity) เงินปันผล (Dividend) และ ส่วนของผู้ถือหุ้นสามัญ (Number of shares) โดยจะพิจารณาโหนดใดเป็นลำดับถัดไปจะพิจารณาจากจำนวนของทุนจดทะเบียนเป็นสำคัญ

ส่วนเทคนิคซัพพอร์ตเวกเตอร์แมชชีนตัวแปรที่ส่งผลกระทบรุนแรง ได้แก่ จำนวนหุ้น, เงินปันผล และกำไร ซึ่งทั้งตัวแปรจำนวนหุ้น เงินปันผล และ กำไรถือเป็นข้อมูลสำคัญที่รายงานในงบการเงิน การรายงานผลของข้อมูลต่าง ๆ ในงบการเงินต้องมีความถูกต้องโปร่งใส เช่น กำไรที่รายงานต้องเป็นกำไรที่มีคุณภาพที่สามารถสะท้อนผลการดำเนินงานที่แท้จริงของกิจการ และสามารถตรวจสอบได้ รวมทั้งใช้พยากรณ์ผลการดำเนินงานและมูลค่าของกิจการในอนาคตได้ด้วย (กนกอร แก้วประภา กิ่งกาญจน์ มูลเมือง และมาลี จตุรัส, 2564)

สำหรับวิธี Generalized Multiple Linear Regression Model ตัวแปรที่มีผลต่อราคาตลาดมากที่สุดแต่เป็นไปในทิศทางตรงกันข้าม คือ ส่วนของผู้ถือหุ้น ส่วนตัวแปรที่ให้ผลในทิศทางตรงข้ามตัวอื่นได้แก่ ทุนจดทะเบียน และ จำนวนหุ้นสามัญ ส่วนตัวแปรที่ส่งผลเป็นบวกได้แก่ ส่วนของการรายได้ กำไรสุทธิ และ เงินปันผลจ่าย ยิ่งตัวแปรเหล่านี้ยิ่งมีค่ามากยิ่งส่งผลให้ราคาตลาดของหุ้นมีค่ามากตามไปด้วย อย่างไรก็ตามในอนาคตถ้าต้องการความถูกต้องแม่นยำในการทำนายมากขึ้นอาจต้องใช้ตัวแปรมากยิ่งขึ้นทั้งปัจจัยภายในและปัจจัยนอกที่เกี่ยวข้อง เช่น ความนิยม ดัชนีตลาด สภาพเศรษฐกิจ สังคมและ การเมืองและเศรษฐกิจ และ ข้อมูลการวิเคราะห์จากนักวิเคราะห์ ข้อมูลจากสื่อโซเชียลต่าง ๆ เป็นต้น ข้อเสนอแนะในการศึกษาวิจัยครั้งถัดไป ควรทำการการศึกษาวิจัยที่ส่งผลต่อราคาตลาดของหุ้นสามัญให้ครอบคลุมทั้งปัจจัยภายในและภายนอกให้มากยิ่งขึ้น แต่ทั้งนี้ก็ต้องมีการเก็บข้อมูลที่เป็นระบบให้ดียิ่งขึ้นและมีจำนวนข้อมูลมากขึ้น รวมถึงเทคนิคที่เลือกใช้ก็ต้องมีความสามารถที่จะพยากรณ์ได้อย่างถูกต้องและแม่นยำมากยิ่งขึ้นต่อไป

เอกสารอ้างอิง

- กนกพร แก้วประภา กิ่งกาญจน์ มูลเมือง และ มาลี จตุรัส. (2564). ความสัมพันธ์ระหว่างเรื่องสำคัญในการตรวจสอบใน รายงานผู้สอบบัญชีและคุณภาพกำไร. *วารสารการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม*, 13(3), 138-153.
- ชินกฤต วงศ์รักษ์. (2564). การตัดสินใจลงทุนหุ้นสามัญในตลาดหลักทรัพย์แห่งประเทศไทยเพื่อวางแผนการเกษียณของ ประชาชนในกรุงเทพมหานคร. *วารสารการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม*, 13(2), 176-190.
- ณัฐฐาภา สันต์งาน และ กิตติพันธ์ คงสวัสดิ์เกียรติ. (2559). ปัจจัยทางเศรษฐกิจที่มีผลต่อราคาหุ้นของบริษัท เด็มโก้ จำกัด (มหาชน). *วารสารวิจัยบัณฑิตศึกษาในกลุ่มธุรกิจและสังคมศาสตร์*, 2(1), 79-91.
- ตลาดหลักทรัพย์แห่งประเทศไทย. (2564). ความเป็นมา บทบาทและหน้าที่. ค้นเมื่อ 18 กุมภาพันธ์ 2564 , จาก https://www.set.or.th/th/about/vision/vision_p1.html
- ประเมษฐ์ ธีรวานนท์ ชัยกร ยิ่งเสรี วรพล พงษ์เพชร และ ธนภัทร ชังคะจิตร. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบ รวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *Journal of information Science and Technology*, 7(1), 12-21
- พิรุณพร ไพฑูรย์เจริญลาภ. (2563). การสร้างแบบจำลองในการพยากรณ์ผลการดำเนินงานจากข้อมูลในงบการเงินของบริษัท ในตลาดหลักทรัพย์กสิกรศึกษา กลุ่มอุตสาหกรรมเกษตรและกลุ่มอุตสาหกรรมอาหาร. *วารสารระบบสารสนเทศ ด้านธุรกิจ (JISB)*, 6(2), 37-48
- ภรณ์ยา ปาลวิสุทธิ. (2559). การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่าง กลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้อในเลือด. *วารสารเทคโนโลยีสารสนเทศ*, 12(1), 54-63
- ยศสยา แสงหิรัญ และ สมชาย เล็กเจริญ. (2561). ปัจจัยที่มีผลต่อการผันแปรของหุ้นกลุ่มเทคโนโลยีโดยใช้เทคนิคการทำเหมืองข้อมูล. การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยรังสิต, ครั้งที่ 13, 16 สิงหาคม 2561, 2582-2590.
- สุภาพรณ คงมณีพรณ. (2562). การหากฎความสัมพันธ์จากฐานข้อมูลการซื้อขายผลิตภัณฑ์เสริมอาหารยี่ห้อมายเฮลท์ของลูกค้า โดยใช้อัลกอริทึมเอพฟิโกรท และการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้ามายเฮลท์โดยเทคนิคคอร์เอพเอ็ม ด้วยโปรแกรมแรพพิดไมเนอร์: กรณีศึกษาของร้านยาเซนแห่งหนึ่ง. *วารสารระบบสารสนเทศด้านธุรกิจ (JISB)*, 4(4), 21-39.
- สุทธิพงศ์ ผ่องแผ้ว. (2561). การศึกษากระบวนการแบ่งกลุ่มผลิตภัณฑ์ด้วยข้อมูลโครงสร้างของเมล็ด. สารนิพนธ์วิทยาศาสตร มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อริพัชร วิจารณ์จิตติคุณ. (2554). ปัจจัยที่มีผลกระทบบต่อราคาหลักทรัพย์ในกลุ่มธนาคารพาณิชย์ในตลาดหลักทรัพย์. สารนิพนธ์เศรษฐศาสตรมหาบัณฑิต สาขาเศรษฐศาสตร์การจัดการ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Harlow, W. V. (2011). Optimal Asset Allocation in Retirement: A Downside Risk Perspective. *Putnam Institute-Retirement*, 6 (June), 1-15.

- Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, 129, 273–285.
- Hur, J., Raj, M., & Riyanto, Y. (2006). Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade. *World Development*, 34(10), 1728-1741.
- Masoud, N. (2017). The impact of stock market performance upon economic growth. *International Journal of Economics and Financial*, 3(4), 788–798.
- Murkute, A., & Tanuja, S. (2015). Forecasting market price of stock using artificial neural network. *International Journal of Computer Applications*, 124(12), 11-15.
- Polanco-Martínez, J. (2019). Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. *Nonlinear Dynamics*, 369–389.