

แบบจำลองการพยากรณ์ผลการดำเนินงานทางการเงินของบริษัทจดทะเบียน ในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม

Financial Performance Forecasting Models for Listed Companies in the Food and Beverage Industry in the Stock Exchange of Thailand

¹นิชพน อัครินทร์ เจริญสุข, ²สุนา สุทธิเกียรติ และ ³รุจิรัตน์ ปาลีพัฒนสกุล

¹Nichaphon Akrinthorn Charoensuk, ²Suna Sudhikiat and ³Rujirat Paleephatsakul

คณะบัญชี มหาวิทยาลัยธนบุรี

Faculty of Business Administration, Thonburi University, Thailand

E-mail: ¹nichapon.akharin.charoensuk@gmail.com, ²sunamaria90@gmail.com, ³rujirat_pal@thonburi-u.ac.th

Received May 5, 2026; Revised June 19, 2026; Accepted July 5, 2026

บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความสัมพันธ์ระหว่างอัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจกับอัตรากำไรสุทธิของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย 2) เปรียบเทียบประสิทธิภาพของแบบจำลอง Multiple Linear Regression และ Random Forest ในการพยากรณ์อัตรากำไรสุทธิ และ 3) วิเคราะห์ตัวแปรที่มีความสำคัญต่อการพยากรณ์ผลการดำเนินงานของบริษัท การวิจัยนี้เป็นการวิจัยเชิงปริมาณ โดยใช้ข้อมูลทุติยภูมิจากงบการเงินของบริษัทจดทะเบียนในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม จำนวน 40 บริษัท ครอบคลุมช่วงปี พ.ศ. 2562–2567 คัดเลือกกลุ่มตัวอย่างแบบเจาะจง และวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน แบบจำลอง Multiple Linear Regression และแบบจำลอง Random Forest โดยประเมินประสิทธิภาพของแบบจำลองด้วยค่า MAE, MSE, RMSE และ R^2 ผลการวิจัยพบว่า ตัวแปรอิสระจำนวน 41 ตัวแปร จากทั้งหมด 86 ตัวแปร มีความสัมพันธ์กับอัตรากำไรสุทธิอย่างมีนัยสำคัญทางสถิติ โดยตัวแปรที่มีความสัมพันธ์สูงสุด คือ อัตรากำไรจ่ายดำเนินงาน ปี 2566 ผลการเปรียบเทียบแบบจำลองพบว่า Random Forest มีประสิทธิภาพในการพยากรณ์สูงกว่า Multiple Linear Regression โดยมีค่า R^2 เท่ากับ 0.776 ค่า MAE เท่ากับ 4.082 และค่า RMSE เท่ากับ 9.569 ขณะที่แบบจำลอง Multiple Linear Regression มีค่า R^2 ตีลบ ตัวแปรที่มีความสำคัญสูงสุดต่อการพยากรณ์ ได้แก่ อัตรากำไรจ่ายดำเนินงาน อัตรากำไรจากการดำเนินงาน และอัตรากำไรสุทธิในปี 2566 องค์ความรู้จากงานวิจัยนี้ชี้ให้เห็นว่า การบูรณาการข้อมูลทางการเงินกับแบบจำลองการเรียนรู้ของเครื่องสามารถใช้เป็นแนวทางในการพยากรณ์ผลการดำเนินงาน และสนับสนุนการตัดสินใจของผู้บริหาร นักลงทุน และนักวิเคราะห์ทางการเงินได้อย่างมีประสิทธิภาพ

คำสำคัญ: อัตราส่วนทางการเงิน; การพยากรณ์ผลการดำเนินงานทางการเงิน; อุตสาหกรรมอาหารและเครื่องดื่ม

Abstract

This article aimed to: 1) study the relationship between financial ratios, internal firm factors, and macroeconomic factors and the net profit margin of companies in the food and beverage industry listed on the Stock Exchange of Thailand; 2) compare the performance of

Multiple Linear Regression and Random Forest models in forecasting net profit margin; and 3) analyze the variables that are important for forecasting firm performance. This study employed a quantitative research design using secondary data from the financial statements of 40 listed companies in the food and beverage industry, covering the period from 2019 to 2024. The sample was selected using purposive sampling. Data were analyzed using descriptive statistics, Spearman's rank correlation coefficient, Multiple Linear Regression, and Random Forest models. Model performance was evaluated using MAE, MSE, RMSE, and R^2 . The findings revealed that 41 independent variables out of a total of 86 variables were significantly associated with net profit margin. The variable with the strongest relationship was the operating expense ratio in 2023. The model comparison results showed that Random Forest outperformed Multiple Linear Regression in forecasting performance, with an R^2 of 0.776, MAE of 4.082, and RMSE of 9.569, whereas the Multiple Linear Regression model yielded a negative R^2 . The most important variables for prediction were the operating expense ratio, operating profit margin, and net profit margin in 2023. The knowledge gained from this study indicates that integrating financial information with machine learning models can serve as an approach for forecasting firm performance and effectively supporting the decision-making of executives, investors, and financial analysts.

Keywords: Financial Ratios; Financial Performance Forecasting; Food and Beverage Industry

บทนำ

อุตสาหกรรมอาหารและเครื่องดื่มถือเป็นกลไกขับเคลื่อนเศรษฐกิจไทยที่มีความสำคัญสูง โดยสำนักงานเศรษฐกิจอุตสาหกรรม (2566) รายงานว่าอุตสาหกรรมนี้มีสัดส่วนในผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ประมาณร้อยละ 10 สร้างการจ้างงานกว่า 1.17 ล้านคน และมีมูลค่าการส่งออกเติบโตเฉลี่ยร้อยละ 5.0–6.0 ต่อปี ด้วยขนาดและความสำคัญดังกล่าว การพยากรณ์ผลการดำเนินงานทางการเงิน โดยเฉพาะอัตรากำไรสุทธิ (Net Profit Margin) จึงเป็นประเด็นที่ได้รับความสนใจทั้งในภาควิชาการและภาคปฏิบัติ งานวิจัยที่ผ่านมาได้พิสูจน์ว่าอัตราส่วนทางการเงินเป็นตัวแปรที่ทรงพลังในการพยากรณ์ผลการดำเนินงานของกิจการ (Ou & Penman, 1989) แบบจำลองการถดถอยพหุคูณ (Multiple Linear Regression: MLR) เป็นแนวทางที่นิยมใช้กันแพร่หลายเนื่องจากอธิบายความสัมพันธ์เชิงเส้นได้ชัดเจนและตีความง่าย ต่อมา Fama & French (2000) และ Piotroski (2000) ชี้ให้เห็นข้อจำกัดของแบบจำลองเชิงเส้นต่อความสัมพันธ์ที่ซับซ้อนในข้อมูลทางการเงิน ขณะที่ Breiman (2001) นำเสนอแบบจำลอง Random Forest ซึ่งเป็น Ensemble Learning ที่สามารถจัดการกับข้อมูลมิติสูงและความสัมพันธ์ไม่เชิงเส้นได้ดีกว่า ต่อมา Gu et al. (2020) เปรียบเทียบแบบจำลอง Machine Learning หลายประเภทในบริบทตลาดหุ้นและพบว่า Random Forest ให้ความแม่นยำสูงกว่า MLR อย่างมีนัยสำคัญ อย่างไรก็ตาม งานวิจัยที่เปรียบเทียบทั้งสองแบบจำลองในบริบทของตลาดหุ้นไทย โดยเฉพาะในระดับอุตสาหกรรมเฉพาะยังมีจำกัด ซึ่งถือเป็นช่องว่างขององค์ความรู้ที่งานวิจัยนี้มุ่งเติมเต็ม

พื้นที่วิจัยของงานนี้คือบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย (SET) กลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม ซึ่งประกอบด้วย 2 หมวดย่อย คือ หมวดอาหาร (Food) และหมวดเครื่องดื่ม (Beverage) รวมกันทั้งสิ้น 48 บริษัท อุตสาหกรรมนี้มีลักษณะเฉพาะที่ทำให้การพยากรณ์ผลการดำเนินงานมีความท้าทาย ได้แก่ ความผันผวนของต้นทุนวัตถุดิบตามฤดูกาลและราคาสินค้าโภคภัณฑ์โลก การแข่งขันสูงทั้งจากผู้ผลิตในประเทศและต่างประเทศ และความอ่อนไหวต่ออัตราเงินเฟ้อและกำลังซื้อของผู้บริโภค (สำนักงานคณะกรรมการ

กำกับหลักทรัพย์และตลาดหลักทรัพย์ [ก.ล.ต.], 2566) ผู้วิจัยได้ทำการสังเกตการณ์และเก็บรวบรวมข้อมูลทุติยภูมิจากรายงานทางการเงินที่เปิดเผยต่อสาธารณะผ่านระบบ SETSMART และเว็บไซต์ตลาดหลักทรัพย์แห่งประเทศไทย ซึ่งครอบคลุมช่วงเวลา 6 ปี (พ.ศ. 2562–2567) สะท้อนทั้งสถานการณ์เศรษฐกิจปกติและช่วงหลังวิกฤตโควิด-19 ที่ส่งผลกระทบต่ออุตสาหกรรมนี้อย่างมีนัยสำคัญ งานวิจัยที่ผ่านมาในพื้นที่ใกล้เคียง เช่น วิศรุต แก้วมหา และวิศิต ปัญญาฉัตรพร (2564) ที่ใช้ Random Forest พยากรณ์ผลตอบแทนหลักทรัพย์ไทย และ ญัฐกร สุขมัน (2563) ที่ศึกษาตัวแปรทางการเงินในกลุ่มอุตสาหกรรมอาหาร ยังไม่ได้เปรียบเทียบแบบจำลองทั้งสองอย่างเป็นระบบบนชุดข้อมูลเดียวกัน

บทความวิจัยนี้นำเสนอโดยแบ่งเนื้อหาออกเป็น 5 ส่วนหลัก ได้แก่ (1) บทนำ ซึ่งอธิบายความสำคัญ ช่องว่างของงานวิจัย และวัตถุประสงค์ (2) การทบทวนวรรณกรรม ครอบคลุมแนวคิดอัตราส่วนทางการเงิน ทฤษฎีพยากรณ์ผลการดำเนินงาน และงานวิจัยที่เกี่ยวข้องกับ MLR และ Random Forest (3) ระเบียบวิธีวิจัย อธิบายกระบวนการคัดกรองกลุ่มตัวอย่าง การเตรียมข้อมูล การคัดเลือกตัวแปร และการสร้างแบบจำลอง (4) ผลการวิจัย นำเสนอผลการวิเคราะห์ความสัมพันธ์ การเปรียบเทียบแบบจำลอง และตัวแปรสำคัญ และ (5) สรุปและข้อเสนอแนะ งานวิจัยนี้คาดว่าจะจะเป็นประโยชน์ต่อวงวิชาการในการขยายองค์ความรู้ด้านการประยุกต์ใช้ Machine Learning กับข้อมูลทางการเงินในบริบทตลาดทุนไทย และต่อภาคปฏิบัติในการช่วยให้นักลงทุนผู้บริหาร และนักวิเคราะห์ทางการเงินสามารถเลือกใช้เครื่องมือพยากรณ์ที่เหมาะสม ลดความเสี่ยงในการตัดสินใจ และเพิ่มประสิทธิภาพการบริหารกิจการในอุตสาหกรรมอาหารและเครื่องดื่มได้อย่างเป็นระบบ

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาความสัมพันธ์ระหว่างอัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ กับอัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง Multiple Linear Regression และ Random Forest ในการพยากรณ์อัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย
3. เพื่อวิเคราะห์และระบุตัวแปรที่มีความสำคัญต่อการพยากรณ์อัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย

สมมติฐานการวิจัย

สมมติฐานที่ 1 อัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับอัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม

สมมติฐานที่ 2 แบบจำลอง Random Forest มีประสิทธิภาพในการพยากรณ์อัตรากำไรสุทธิ (Net Profit Margin) สูงกว่าแบบจำลอง Multiple Linear Regression อย่างมีนัยสำคัญทางสถิติ

การทบทวนวรรณกรรม

จากการสืบค้นเอกสาร บทความ และงานวิจัยที่เกี่ยวข้องกับการพยากรณ์ผลการดำเนินงานทางการเงินด้วยอัตราส่วนทางการเงินและ Machine Learning พบว่างานวิจัยส่วนใหญ่ในต่างประเทศ เช่น Gu, Kelly, and Xiu (2020) และ Wasserbacher and Spindler (2022) มุ่งเน้นการพิสูจน์ศักยภาพของ Machine Learning ใน

ระดับตลาดทุนหรือกลุ่มอุตสาหกรรมขนาดใหญ่ ในขณะที่งานวิจัยในบริบทประเทศไทย เช่น นริศรา ประไพ และคณะ (2565) และ วิศรุต แก้วมหา และวิศิต ปัญญาฉัตรพร (2564) ยังจำกัดอยู่ที่การพยากรณ์ผลตอบแทนหลักทรัพย์เป็นหลัก มากกว่าการพยากรณ์ผลการดำเนินงานเชิงบัญชี-การเงินในระดับกิจการ นอกจากนี้ Hezam, Luong, and Anthonysamy (2025) ยังชี้ว่าประสิทธิภาพของแบบจำลอง Machine Learning ขึ้นอยู่กับบริบทอุตสาหกรรมและลักษณะข้อมูลเฉพาะกลุ่ม จึงยังไม่พบงานวิจัยที่น่าอัตราส่วนทางการเงิน ปัจจัยภายในกิจการ และปัจจัยมหภาคมาประยุกต์ใช้ร่วมกับ Random Forest เพื่อพยากรณ์อัตรากำไรสุทธิของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทยโดยตรง ซึ่งถือเป็นช่องว่างของความรู้ (Knowledge Gap) ที่งานวิจัยนี้มุ่งเติมเต็ม

การพยากรณ์ผลการดำเนินงานของกิจการ (Business Performance Forecasting) เป็นเครื่องมือสำคัญในการสนับสนุนการตัดสินใจเชิงกลยุทธ์ขององค์กร โดยอาศัยข้อมูลในอดีตและปัจจุบันเพื่อคาดการณ์ผลลัพธ์ทางการเงินในอนาคต โดยเฉพาะอัตรากำไรสุทธิ (Net Profit Margin) ซึ่งเป็นตัวชี้วัดสำคัญของความสามารถในการทำกำไรและประสิทธิภาพในการดำเนินงานของกิจการ (Fama & French, 2000) ในบริบทของการวิเคราะห์ทางการเงิน อัตราส่วนทางการเงิน (Financial Ratios) ถูกนำมาใช้อย่างแพร่หลายในการสะท้อนสถานะทางการเงินขององค์กรในหลายมิติ ได้แก่ ความสามารถในการทำกำไร สภาพคล่อง ประสิทธิภาพในการดำเนินงาน และโครงสร้างเงินทุน โดยงานของ Ou and Penman (1989) ชี้ให้เห็นว่าอัตราส่วนทางการเงินมีความสามารถในการอธิบายและพยากรณ์ผลการดำเนินงานของกิจการได้อย่างมีนัยสำคัญ นอกจากนี้ อัตราส่วนทางการเงินยังมีบทบาทสำคัญในการสะท้อนสัญญาณเชิงเศรษฐกิจของกิจการ ทั้งด้านความสามารถในการสร้างกำไร ความเสี่ยงทางการเงิน และประสิทธิภาพในการบริหารทรัพยากร ซึ่งเป็นข้อมูลพื้นฐานที่นักลงทุน ผู้บริหาร และนักวิเคราะห์ใช้ประกอบการประเมินผลการดำเนินงานในอนาคต

ในด้านระเบียบวิธีการวิเคราะห์ แบบจำลองการถดถอยพหุคูณ (Multiple Linear Regression: MLR) เป็นวิธีการทางสถิติที่ได้รับความนิยมในการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม เนื่องจากมีข้อดีในด้านความโปร่งใสและสามารถตีความผลลัพธ์ได้อย่างชัดเจน อย่างไรก็ตาม แบบจำลองดังกล่าวมีข้อจำกัดในการจัดการกับข้อมูลที่มีความซับซ้อนหรือมีความสัมพันธ์แบบไม่เป็นเชิงเส้น รวมถึงกรณีที่ข้อมูลมีลักษณะมิติสูง (high-dimensional data) ซึ่งจำนวนตัวแปรอิสระมีมากกว่าจำนวนตัวอย่าง ($p > n$) อันส่งผลให้เกิดปัญหา overfitting และลดความแม่นยำของแบบจำลอง (Gu et al., 2020) โดยเฉพาะข้อมูลทางการเงินของบริษัทจดทะเบียนที่มักประกอบด้วยตัวแปรจำนวนมากและมีความสัมพันธ์กันเองสูง เช่น อัตราส่วนกำไร สภาพคล่อง โครงสร้างเงินทุน และประสิทธิภาพการดำเนินงาน การใช้แบบจำลองเชิงเส้นเพียงอย่างเดียวอาจไม่สามารถจับความสัมพันธ์เชิงซ้อนหรือปฏิสัมพันธ์ระหว่างตัวแปรได้ครบถ้วน

ข้อจำกัดดังกล่าวส่งผลให้เกิดการพัฒนาและนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เข้ามาประยุกต์ใช้ในการพยากรณ์ผลการดำเนินงานทางการเงินมากขึ้น ทั้งนี้ Wasserbacher and Spindler (2022) ชี้ให้เห็นว่า Machine Learning มีความเหมาะสมกับงานพยากรณ์ทางการเงิน เนื่องจากสามารถประมวลผลข้อมูลจำนวนมากและค้นหารูปแบบความสัมพันธ์ในข้อมูลได้อย่างยืดหยุ่น อย่างไรก็ตาม การใช้ Machine Learning ควรระมัดระวังในการตีความผลลัพธ์ เนื่องจากแบบจำลองมุ่งเน้นความแม่นยำในการพยากรณ์มากกว่าการอธิบายความสัมพันธ์เชิงเหตุและผล

หนึ่งในเทคนิคที่ได้รับความนิยมคือแบบจำลอง Random Forest ซึ่งเป็นวิธีการเรียนรู้แบบ Ensemble Learning ที่พัฒนาโดย Breiman (2001) โดยใช้การรวมผลของต้นไม้ตัดสินใจจำนวนมากเพื่อเพิ่มความแม่นยำในการพยากรณ์ แบบจำลองนี้มีจุดเด่นในการจัดการกับความสัมพันธ์ที่ไม่เป็นเชิงเส้นและลดปัญหา overfitting ได้อย่างมีประสิทธิภาพ (นริศรา ประไพ และคณะ, 2565) อีกทั้งยังสามารถประเมินความสำคัญของตัวแปร

(Feature Importance) ได้ ซึ่งเป็นประโยชน์ต่อการตีความผลลัพธ์ในเชิงการวิเคราะห์ทางการเงิน คุณลักษณะดังกล่าวทำให้ Random Forest เหมาะสมกับข้อมูลทางการเงินที่มีตัวแปรจำนวนมากและมีความสัมพันธ์ซับซ้อน เนื่องจากสามารถระบุรูปแบบความสัมพันธ์ที่ไม่จำเป็นต้องกำหนดสมมติฐานเชิงเส้นล่วงหน้า งานวิจัยในบริบทประเทศไทยพบว่าการใช้เทคนิค Machine Learning เช่น Random Forest สามารถเพิ่มความแม่นยำในการพยากรณ์ผลตอบแทนของหลักทรัพย์ได้ดีกว่าแบบจำลองเชิงสถิติแบบดั้งเดิม โดยมีค่าความคลาดเคลื่อนต่ำกว่า และสามารถรองรับข้อมูลที่มีความซับซ้อนได้มากกว่า (วิศรุต แก้วมหา และวิศรุต ปัญญาฉัตรพร, 2564) สอดคล้องกับ Gu et al. (2020) ที่พบว่าแบบจำลองกลุ่ม Tree-Based Methods และ Neural Networks สามารถสร้างผลการพยากรณ์ที่ดีกว่าแนวทางเชิงเส้นในบริบทข้อมูลทางการเงิน เนื่องจากสามารถรองรับความสัมพันธ์แบบไม่เป็นเชิงเส้นและปฏิสัมพันธ์ระหว่างตัวแปรได้ดีกว่า

ในช่วงหลัง งานวิจัยด้านการเงินเริ่มให้ความสำคัญกับการใช้ Machine Learning เพื่อพยากรณ์ผลการดำเนินงานของกิจการมากขึ้น Tutcu et al. (2024) ศึกษาการพยากรณ์ผลการดำเนินงานทางการเงินของบริษัทในอุตสาหกรรมเทคโนโลยีสารสนเทศ โดยใช้ตัวชี้วัด ROA และ ROE ร่วมกับแบบจำลอง Machine Learning และพบว่าแนวทางดังกล่าวสามารถใช้สนับสนุนการประเมินผลการดำเนินงานของกิจการได้ ขณะที่ Hezam, Luong, and Anthonysamy (2025) ได้สังเคราะห์งานวิจัยเกี่ยวกับ Machine Learning ในการพยากรณ์ผลการดำเนินงานของกิจการ และชี้ให้เห็นว่าแบบจำลอง Machine Learning ถูกนำมาใช้เพิ่มขึ้นในงานด้าน Predictive Analytics และ Business Intelligence โดยเฉพาะเมื่อต้องจัดการกับข้อมูลจำนวนมากและตัวแปรหลายมิติ อย่างไรก็ตาม งานวิจัยเหล่านี้ยังสะท้อนให้เห็นว่า บริบทของอุตสาหกรรม ประเทศ และลักษณะข้อมูลที่ใช้วิเคราะห์ยังมีผลต่อประสิทธิภาพของแบบจำลอง ดังนั้น การศึกษาเฉพาะกลุ่มอุตสาหกรรมจึงยังมีความจำเป็นเพื่อให้ได้ข้อค้นพบที่สอดคล้องกับบริบทของข้อมูลจริง

เพื่อให้เห็นความสัมพันธ์และพัฒนาการของวรรณกรรมที่ทบทวนอย่างเป็นระบบ สามารถเรียงลำดับวรรณกรรมตามประเด็นที่เกี่ยวข้องได้ดังนี้

1. Ou and Penman (1989) และ Fama and French (2000) วางรากฐานแนวคิดว่าอัตราส่วนทางการเงินและความสามารถในการทำกำไรในอดีตสามารถใช้พยากรณ์ผลการดำเนินงานในอนาคตได้
2. Gu, Kelly, and Xiu (2020) ชี้ให้เห็นข้อจำกัดของแบบจำลองเชิงเส้นแบบดั้งเดิมในการจัดการข้อมูลมิติสูง อันเป็นจุดเปลี่ยนสำคัญที่ผลักดันให้เกิดการนำ Machine Learning เข้ามาแทนที่หรือเสริมแบบจำลองเชิงเส้น
3. Wasserbacher and Spindler (2022) และ Breiman (2001) อธิบายหลักการและจุดเด่นของ Machine Learning และ Random Forest โดยเฉพาะ ในการจับความสัมพันธ์แบบไม่เป็นเชิงเส้นและลดปัญหา overfitting ซึ่งเป็นพื้นฐานเชิงทฤษฎีของแบบจำลองที่งานวิจัยนี้นำมาใช้
4. นริศรา ประไพ และคณะ (2565) และ วิศรุต แก้วมหา และวิศรุต ปัญญาฉัตรพร (2564) นำเสนอหลักฐานเชิงประจักษ์ในบริบทประเทศไทยว่า Random Forest ให้ผลพยากรณ์ที่ดีกว่าแบบจำลองสถิติดั้งเดิมในตลาดหุ้นไทย
5. Tutcu et al. (2024) และ Hezam, Luong, and Anthonysamy (2025) สะท้อนแนวโน้มล่าสุดของการประยุกต์ใช้ Machine Learning ในการพยากรณ์ผลการดำเนินงานของกิจการในต่างประเทศ พร้อมทั้งชี้ว่าประสิทธิภาพของแบบจำลองยังขึ้นอยู่กับบริบทอุตสาหกรรมเฉพาะ ซึ่งเป็นที่มาของช่องว่างความรู้ที่งานวิจัยนี้มุ่งศึกษาในบริบทอุตสาหกรรมอาหารและเครื่องดื่มของไทย

โดยสรุป การทบทวนวรรณกรรมแสดงให้เห็นว่าผู้วิจัยได้นำแนวคิดของ Ou and Penman (1989) และ Fama and French (2000) มาใช้เป็นกรอบในการคัดเลือกอัตราส่วนทางการเงินเป็นตัวแปรอิสระของการวิจัย นำ

ข้อจำกัดของแบบจำลองเชิงเส้นตามที่ Gu, Kelly, and Xiu (2020) ชี้ไว้ มาเป็นเหตุผลในการเลือกใช้ Random Forest ตามแนวคิดของ Breiman (2001) เป็นแบบจำลองเปรียบเทียบกับ Multiple Linear Regression และอ้างอิงผลการศึกษาของ นริศรา ประไพ และคณะ (2565) และ วิศรุต แก้วมหา และวิริศ ปัญญาฉัตรพร (2564) เพื่อสนับสนุนสมมติฐานว่า Random Forest จะให้ผลพยากรณ์ที่ดีกว่าในบริบทข้อมูลการเงินไทย นอกจากนี้ ข้อเสนอของ Hezam, Luong, and Anthonysamy (2025) เกี่ยวกับความจำเพาะเชิงอุตสาหกรรมได้ถูกนำมาใช้ กำหนดขอบเขตการวิจัยให้มุ่งเน้นเฉพาะกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย เพื่อเติมเต็มช่องว่างความรู้ที่ยังไม่มีการศึกษาในบริบทดังกล่าวมาก่อน

ตารางที่ 1 การสังเคราะห์วรรณกรรมที่เกี่ยวข้องกับการพยากรณ์ผลการดำเนินงานทางการเงิน

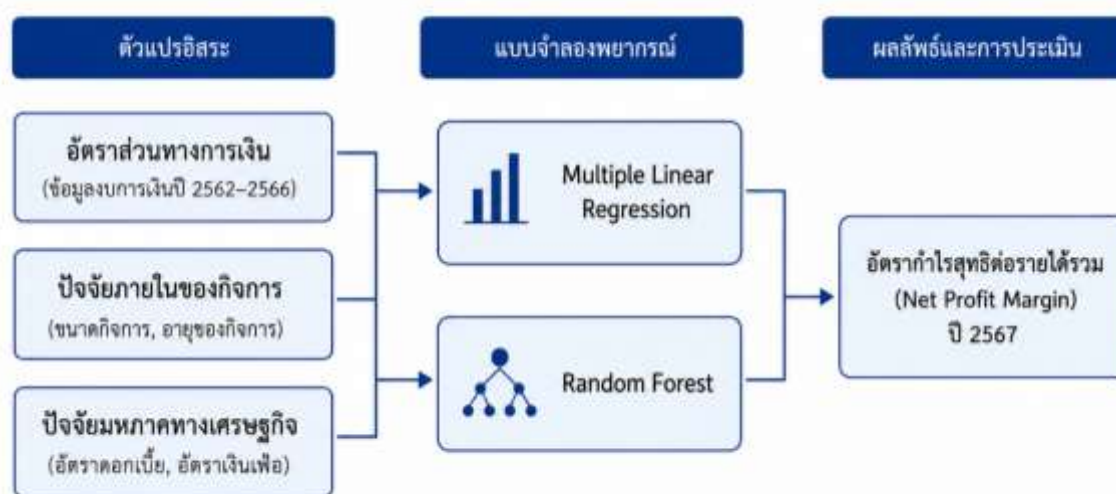
ผู้วิจัย/ปี	ประเด็นศึกษา	วิธีการ/แบบจำลอง	ข้อค้นพบและช่องว่างที่เกี่ยวข้อง
Ou and Penman (1989)	อัตราส่วนทางการเงินกับการพยากรณ์ผลการดำเนินงาน	การวิเคราะห์อัตราส่วนทางการเงิน	อัตราส่วนทางการเงินสามารถใช้พยากรณ์ผลการดำเนินงานได้ แต่ยังไม่ได้เปรียบเทียบกับแบบจำลอง Machine Learning
Breiman (2001)	การพัฒนาแบบจำลอง Random Forest	Random Forest / Ensemble Learning	Random Forest รองรับการสัมพันธ์ซับซ้อนและลดปัญหา Overfitting ได้ แต่ยังคงประยุกต์ใช้กับข้อมูลทางการเงิน
Gu et al. (2020)	การประยุกต์ Machine Learning กับข้อมูลทางการเงิน	Tree-based Methods และ Neural Networks	Machine Learning ให้ผลพยากรณ์ดีกว่าแบบจำลองเชิงเส้นในข้อมูลการเงินที่ซับซ้อน แต่ยังไม่ได้นับอุตสาหกรรมเฉพาะในตลาดทุนไทย
Wasserbacher and Spindler (2022)	Machine Learning ในการพยากรณ์ทางการเงิน	Machine Learning for Financial Forecasting	Machine Learning เหมาะกับข้อมูลจำนวนมากและหลายมิติ แต่ต้องระมัดระวังการตีความผลลัพธ์เชิงเหตุและผล
วิศรุต แก้วมหา และวิริศ ปัญญาฉัตรพร (2564)	การพยากรณ์ผลตอบแทนหลักทรัพย์ในประเทศไทย	Machine Learning / Random Forest	Machine Learning ให้ความแม่นยำสูงกว่าแบบจำลองสถิติดั้งเดิม แต่ยังไม่ได้ศึกษา Net Profit Margin ของกลุ่มอาหารและเครื่องดื่ม
Tutcu et al. (2024); Hezam et al. (2025)	การใช้ Machine Learning เพื่อพยากรณ์ผลการดำเนินงานของกิจการ	Machine Learning / Predictive Analytics	งานวิจัยล่าสุดยืนยันแนวโน้มการใช้ Machine Learning ในการประเมินและพยากรณ์ผลการดำเนินงาน แต่ผลลัพธ์ยังขึ้นอยู่กับบริบทอุตสาหกรรม ประเทศ และลักษณะข้อมูล

จากการสังเคราะห์วรรณกรรมในตารางที่ 1 พบว่า งานวิจัยที่ผ่านมาให้ความสำคัญกับการใช้อัตราส่วนทางการเงินและแบบจำลอง Machine Learning ในการพยากรณ์ผลลัพธ์ทางการเงินเพิ่มขึ้น อย่างไรก็ตาม ยังมีข้อจำกัดในด้านการเปรียบเทียบประสิทธิภาพของแบบจำลองเชิงสถิติและแบบจำลอง Machine Learning ภายใต้อข้อมูลชุดเดียวกัน โดยเฉพาะในบริบทของกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มในตลาดหลักทรัพย์แห่งประเทศไทย

ไทย อีกทั้งยังขาดการวิเคราะห์ว่าตัวแปรทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจใดมีความสำคัญต่อการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวม ดังนั้น งานวิจัยนี้จึงมุ่งเติมเต็มช่องว่างดังกล่าวโดยเปรียบเทียบประสิทธิภาพของแบบจำลอง Multiple Linear Regression และ Random Forest พร้อมทั้งวิเคราะห์ตัวแปรสำคัญที่มีผลต่อการพยากรณ์ผลการดำเนินงานของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม

กรอบแนวคิดการวิจัย

กรอบแนวคิดของงานวิจัยนี้มุ่งศึกษาการพยากรณ์ผลการดำเนินงานทางการเงินของบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย กลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม โดยกำหนดตัวแปรอิสระเป็น 3 กลุ่ม ได้แก่ อัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ ซึ่งใช้ข้อมูลย้อนหลังในช่วงปี พ.ศ. 2562–2566 เป็นข้อมูลนำเข้าในการพยากรณ์ จากนั้นนำข้อมูลดังกล่าวมาวิเคราะห์ผ่านแบบจำลอง Multiple Linear Regression และ Random Forest เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวม (Net Profit Margin) ของปี พ.ศ. 2567 ซึ่งเป็นตัวแปรตามของการวิจัย โดยพิจารณาจากค่าความคลาดเคลื่อนและความสามารถในการอธิบายของแบบจำลอง ได้แก่ MAE, MSE, RMSE และ R^2 รวมทั้งวิเคราะห์ความสำคัญของตัวแปรจากค่า Feature Importance ของแบบจำลอง Random Forest เพื่อระบุตัวแปรที่มีบทบาทสำคัญต่อการพยากรณ์ผลการดำเนินงานทางการเงิน



ภาพที่ 1 กรอบแนวคิดการพยากรณ์ผลการดำเนินงานทางการเงินของบริษัทจดทะเบียนในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม

ระเบียบวิธีวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงปริมาณ โดยใช้ข้อมูลทุติยภูมิจากงบการเงินของบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย กลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม ระหว่างปี พ.ศ. 2562–2567 เพื่อวิเคราะห์ความสัมพันธ์ของตัวแปรทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ รวมทั้งเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ผลการดำเนินงานทางการเงิน โดยใช้ข้อมูลปี พ.ศ. 2562–2566 เป็นข้อมูลนำเข้าในการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวม (Net Profit Margin) ของปี พ.ศ. 2567 การวิจัยแบ่งออกเป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การกำหนดประชากร กลุ่มตัวอย่าง และการรวบรวมข้อมูล ประชากรที่ใช้ในการวิจัย คือ บริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย กลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม กลุ่มตัวอย่าง คัดเลือกด้วยวิธีแบบเจาะจง โดยมีเกณฑ์การคัดเลือก คือ 1) เป็นบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม 2) มีงบการเงินครบถ้วนต่อเนื่องระหว่างปี พ.ศ. 2562–2567 3) ไม่อยู่ระหว่างการฟื้นฟูกิจการ และ 4) มีข้อมูลครบถ้วนสำหรับการคำนวณตัวแปรที่ใช้ในการวิจัย ได้กลุ่มตัวอย่างจำนวน 40 บริษัท หน่วยวิเคราะห์ของงานวิจัย คือ บริษัทจดทะเบียนในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม โดยข้อมูลทางการเงินถูกจัดเก็บจากงบการเงินประจำปีและฐานข้อมูลของตลาดหลักทรัพย์แห่งประเทศไทย ส่วนข้อมูลปัจจัยมหภาคทางเศรษฐกิจ เช่น อัตราดอกเบี้ยและอัตราเงินเฟ้อ รวบรวมจากแหล่งข้อมูลเศรษฐกิจมหภาคที่เกี่ยวข้อง เครื่องมือที่ใช้ในการวิจัย คือ แบบบันทึกข้อมูลทุติยภูมิ ซึ่งใช้ในการรวบรวมและจัดระบบข้อมูลจากงบการเงิน ฐานข้อมูลตลาดหลักทรัพย์แห่งประเทศไทย และข้อมูลเศรษฐกิจมหภาคที่เกี่ยวข้อง

ขั้นตอนที่ 2 การคำนวณตัวแปรและการเตรียมข้อมูล ผู้วิจัยนำข้อมูลจากงบการเงินของบริษัทกลุ่มตัวอย่างมาคำนวณเป็นตัวแปรที่ใช้ในการวิจัย โดยตัวแปรตาม คือ อัตรากำไรสุทธิ (Net Profit Margin) ของปี พ.ศ. 2567 ส่วนตัวแปรอิสระประกอบด้วย 3 กลุ่ม ได้แก่ 1) อัตราส่วนทางการเงินจากข้อมูลปี พ.ศ. 2562–2566 2) ปัจจัยภายในของกิจการ เช่น ขนาดและอายุของกิจการ และ 3) ปัจจัยมหภาคทางเศรษฐกิจ เช่น อัตราดอกเบี้ยและอัตราเงินเฟ้อ รวมตัวแปรที่ใช้ในการวิเคราะห์ทั้งสิ้น 86 ตัวแปร ก่อนนำข้อมูลเข้าสู่การวิเคราะห์ ผู้วิจัยตรวจสอบความครบถ้วน ความถูกต้อง และความสอดคล้องของข้อมูล รวมถึงพิจารณาค่าผิดปกติของข้อมูลทางการเงิน เพื่อให้ชุดข้อมูลมีความเหมาะสมสำหรับการวิเคราะห์เชิงสถิติและการสร้างแบบจำลองพยากรณ์ ทั้งนี้ การจัดเตรียมข้อมูลดังกล่าวมีวัตถุประสงค์เพื่อลดความคลาดเคลื่อนที่อาจเกิดจากข้อมูลไม่ครบถ้วนหรือค่าที่ผิดปกติ และเพื่อให้ผลการวิเคราะห์มีความน่าเชื่อถือมากยิ่งขึ้น

ขั้นตอนที่ 3 การวิเคราะห์และเปรียบเทียบแบบจำลอง ผู้วิจัยวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา ได้แก่ ค่าเฉลี่ย มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด และค่าต่ำสุด เพื่ออธิบายลักษณะเบื้องต้นของข้อมูล และวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระกับอัตรากำไรสุทธิต่อรายได้รวมด้วยค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman Rank Correlation) เนื่องจากข้อมูลทางการเงินอาจมีการกระจายไม่เป็นปกติและมีค่าผิดปกติบางรายการ จากนั้นผู้วิจัยสร้างแบบจำลองการพยากรณ์ 2 ประเภท ได้แก่ แบบจำลอง Multiple Linear Regression และแบบจำลอง Random Forest โดยแบ่งข้อมูลเป็นชุดฝึกสอนและชุดทดสอบในอัตราส่วน 80:20 เพื่อใช้ชุดข้อมูลส่วนหนึ่งในการสร้างแบบจำลอง และใช้ข้อมูลอีกส่วนหนึ่งในการทดสอบความสามารถในการพยากรณ์ของแบบจำลองกับข้อมูล

สำหรับแบบจำลอง Multiple Linear Regression ผู้วิจัยใช้ Ridge Regression เพื่อลดผลกระทบจากปัญหาความสัมพันธ์สูงระหว่างตัวแปรอิสระและลดความเสี่ยงของการเกิด overfitting ซึ่งอาจเกิดขึ้นจากจำนวนตัวแปรอิสระที่ค่อนข้างมากเมื่อเทียบกับจำนวนกลุ่มตัวอย่าง ส่วนแบบจำลอง Random Forest กำหนดจำนวนต้นไม้ตัดสินใจเท่ากับ 200 ต้น และกำหนดค่า random_state = 42 เพื่อให้ผลการวิเคราะห์สามารถทำซ้ำได้ นอกจากนี้ แบบจำลอง Random Forest ยังถูกใช้ในการประเมินความสำคัญของตัวแปรผ่านค่า Feature Importance เพื่อระบุตัวแปรที่มีบทบาทสำคัญต่อการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวม

การประเมินประสิทธิภาพของแบบจำลองใช้ค่า Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) และค่าสัมประสิทธิ์การกำหนด (R^2) โดยค่า MAE, MSE และ RMSE ใช้สะท้อนระดับความคลาดเคลื่อนของการพยากรณ์ ขณะที่ค่า R^2 ใช้พิจารณาความสามารถของแบบจำลองในการอธิบายความแปรปรวนของตัวแปรตาม นอกจากนี้ ผู้วิจัยเปรียบเทียบความแตกต่างของประสิทธิภาพการพยากรณ์ระหว่างแบบจำลองด้วย Wilcoxon Signed-Rank Test และ Diebold-Mariano Test เพื่อประเมินว่าแบบจำลอง

Random Forest และ Multiple Linear Regression มีความแตกต่างกันในด้านความแม่นยำของการพยากรณ์อย่างมีนัยสำคัญทางสถิติหรือไม่

ผลการวิจัย

การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระ ได้แก่ อัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ กับอัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม ดำเนินการโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman Rank Correlation) เนื่องจากข้อมูลทางการเงินมักมีการกระจายไม่สม่ำเสมอและอาจมีค่าผิดปกติ การใช้สหสัมพันธ์แบบสเปียร์แมนจึงเหมาะสมต่อการวิเคราะห์ความสัมพันธ์เชิงอันดับระหว่างตัวแปร

ผลการวิเคราะห์พบว่า จากตัวแปรอิสระทั้งหมด 86 ตัวแปร มีตัวแปรจำนวน 41 ตัวแปร หรือคิดเป็นร้อยละ 47.7 ที่มีความสัมพันธ์กับอัตรากำไรสุทธิ (Net Profit Margin) อย่างมีนัยสำคัญทางสถิติ โดยตัวแปรที่มีความสัมพันธ์สูงสุด 3 อันดับแรก ได้แก่ อัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม ปี 2566 (M_2566) ซึ่งมีความสัมพันธ์เชิงลบในระดับค่อนข้างสูงกับอัตรากำไรสุทธิ ($\rho = -0.7008$, $p < 0.001$) รองลงมา คือ อัตรากำไรจากการดำเนินงานต่อรายได้รวม ปี 2566 (G_2566) ซึ่งมีความสัมพันธ์เชิงบวกกับอัตรากำไรสุทธิ ($\rho = 0.6544$, $p < 0.001$) และอัตรากำไรจากการดำเนินงานต่อรายได้รวม ปี 2565 (G_2565) ซึ่งมีความสัมพันธ์เชิงบวกเช่นเดียวกัน ($\rho = 0.6503$, $p < 0.001$) ดังแสดงในตารางที่ 2

ตารางที่ 2 ความสัมพันธ์ระหว่างตัวแปรอิสระกับ Net Profit Margin 10 อันดับแรก

ลำดับ	ตัวแปร	ความหมาย	Spearman ρ	p-value
1	M_2566	ค่าใช้จ่ายดำเนินงาน/รายได้รวม ปี 2566	-0.7008	< 0.001***
2	G_2566	กำไรจากการดำเนินงาน/รายได้รวม ปี 2566	0.6544	< 0.001***
3	G_2565	กำไรจากการดำเนินงาน/รายได้รวม ปี 2565	0.6503	< 0.001***
4	M_2565	ค่าใช้จ่ายดำเนินงาน/รายได้รวม ปี 2565	-0.6475	< 0.001***
5	Y_2565	กำไรสุทธิ/รายได้รวม ปี 2565	0.6475	< 0.001***
6	Y_2566	กำไรสุทธิ/รายได้รวม ปี 2566	0.6447	< 0.001***
7	M_2564	ค่าใช้จ่ายดำเนินงาน/รายได้รวม ปี 2564	-0.5814	< 0.001***
8	M_2563	ค่าใช้จ่ายดำเนินงาน/รายได้รวม ปี 2563	-0.5728	< 0.001***
9	Y_2563	กำไรสุทธิ/รายได้รวม ปี 2563	0.5711	< 0.001***
10	Y_2564	กำไรสุทธิ/รายได้รวม ปี 2564	0.5666	< 0.001***

หมายเหตุ: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

จากตารางที่ 2 พบว่า ตัวแปรที่มีความสัมพันธ์กับอัตรากำไรสุทธิสูงสุด คือ อัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม ปี 2566 (M_2566) โดยมีความสัมพันธ์เชิงลบกับอัตรากำไรสุทธิ แสดงให้เห็นว่า หากบริษัทมีสัดส่วนค่าใช้จ่ายดำเนินงานต่อรายได้รวมสูงขึ้น อัตรากำไรสุทธิมีแนวโน้มลดลง สะท้อนถึงความสำคัญของการควบคุมต้นทุนและค่าใช้จ่ายในการดำเนินงานต่อความสามารถในการทำกำไรของบริษัท

ขณะเดียวกัน ตัวแปรกลุ่มอัตรากำไรจากการดำเนินงานต่อรายได้รวม ได้แก่ G_2566 และ G_2565 มีความสัมพันธ์เชิงบวกกับอัตรากำไรสุทธิ แสดงให้เห็นว่า บริษัทที่มีประสิทธิภาพในการสร้างกำไรจากการดำเนินงานสูงในปีก่อนหน้า มีแนวโน้มที่จะมีอัตรากำไรสุทธิสูงขึ้นในปีถัดไป นอกจากนี้ ตัวแปรอัตรากำไรสุทธิต่อ

รายได้รวมในอดีต ได้แก่ Y_2565, Y_2566, Y_2563 และ Y_2564 ก็มีความสัมพันธ์เชิงบวกกับอัตรากำไรสุทธิในปี 2567 ซึ่งสะท้อนลักษณะความต่อเนื่องของผลการดำเนินงานทางการเงินของบริษัท

ดังนั้น ผลการวิจัยตามวัตถุประสงค์ที่ 1 ตัวแปรที่เกี่ยวข้องกับการควบคุมค่าใช้จ่าย การสร้างกำไรจากการดำเนินงาน และผลการดำเนินงานในอดีต เป็นกลุ่มตัวแปรสำคัญที่มีความสัมพันธ์กับอัตรากำไรสุทธิของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มอย่างมีนัยสำคัญทางสถิติ ซึ่งสามารถนำไปใช้เป็นข้อมูลพื้นฐานในการสร้างแบบจำลองพยากรณ์ผลการดำเนินงานในขั้นตอนต่อไปได้

การเปรียบเทียบประสิทธิภาพของแบบจำลอง Multiple Linear Regression และ Random Forest เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์อัตรากำไรสุทธิ พบว่า แบบจำลอง Random Forest มีประสิทธิภาพในการพยากรณ์สูงกว่าแบบจำลอง Multiple Linear Regression อย่างชัดเจน โดยแบบจำลอง RF Baseline ให้ค่าความสามารถในการอธิบาย หรือ R^2 เท่ากับ 0.776 มีค่า MAE เท่ากับ 4.082 และค่า RMSE เท่ากับ 9.569 ขณะที่แบบจำลอง MLR Baseline มีค่า R^2 เท่ากับ -0.809 ค่า MAE เท่ากับ 9.802 และค่า RMSE เท่ากับ 27.174 ดังแสดงในตารางที่ 3

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของแบบจำลอง Multiple Linear Regression และ Random Forest

แบบจำลอง	RMSE	MAE	R^2	หมายเหตุ
RF Baseline	9.569	4.082	0.776	Random Forest
RF Group Model	9.577	4.466	0.775	Random Forest
MLR Baseline	27.174	9.802	-0.809	Ridge $\alpha=10$
MLR Group Model	39.478	10.787	-2.818	Logit+MLR

จากตารางที่ 3 แสดงให้เห็นว่า แบบจำลอง Random Forest สามารถพยากรณ์อัตรากำไรสุทธิได้แม่นยำกว่าแบบจำลอง Multiple Linear Regression ทั้งในด้านค่าความคลาดเคลื่อนที่ต่ำกว่า และค่าความสามารถในการอธิบายที่สูงกว่า โดยค่า MAE และ RMSE ที่ต่ำกว่าสะท้อนว่า Random Forest มีความคลาดเคลื่อนจากค่าจริงน้อยกว่า ในขณะที่ค่า R^2 ที่สูงกว่าแสดงถึงความสามารถในการอธิบายความแปรปรวนของอัตรากำไรสุทธิได้ดีกว่า

สำหรับแบบจำลอง Multiple Linear Regression พบว่าค่า R^2 มีค่าเป็นลบ ซึ่งสะท้อนว่าแบบจำลองมีประสิทธิภาพในการพยากรณ์ต่ำกว่าค่าเฉลี่ยของตัวแปรตาม สาเหตุอาจเกิดจากลักษณะของข้อมูลที่มีจำนวนตัวแปรอิสระสูงเมื่อเทียบกับจำนวนตัวอย่าง กล่าวคือ มีตัวแปรที่ใช้ในการวิเคราะห์ 86 ตัวแปร ขณะที่กลุ่มตัวอย่างจำนวน 40 บริษัท ส่งผลให้เกิดข้อจำกัดด้านโครงสร้างข้อมูลและความเสี่ยงต่อปัญหา Overfitting แม้ว่าจะใช้ Ridge Regression เพื่อลดปัญหาความสัมพันธ์สูงระหว่างตัวแปรอิสระแล้วก็ตาม

ดังนั้น ผลการวิจัยตามวัตถุประสงค์ที่ 2 สรุปได้ว่า แบบจำลอง Random Forest มีความเหมาะสมมากกว่าแบบจำลอง Multiple Linear Regression ในการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวมของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม เนื่องจากให้ค่าความคลาดเคลื่อนต่ำกว่าและมีความสามารถในการอธิบายผลการดำเนินงานสูงกว่า

การวิเคราะห์ตัวแปรสำคัญจากแบบจำลอง Random Forest ผลการวิเคราะห์ความสำคัญของตัวแปรจากแบบจำลอง Random Forest พบว่า ตัวแปรที่มีอิทธิพลต่อการพยากรณ์สูงสุด คือ อัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม ปี 2566 (M_2566) มีค่า Feature Importance เท่ากับร้อยละ 27.54 รองลงมา คือ อัตรากำไรจากการดำเนินงานต่อรายได้รวม ปี 2566 (G_2566) เท่ากับร้อยละ 18.79 และอัตรากำไรสุทธิต่อรายได้รวม ปี

2566 (Y_2566) เท่ากับร้อยละ 12.49 โดยตัวแปรทั้งสามรวมกันมีสัดส่วนความสำคัญร้อยละ 59.92 ของความสำคัญทั้งหมด

ผลดังกล่าวแสดงให้เห็นว่า ตัวแปรด้านการควบคุมค่าใช้จ่ายและความสามารถในการทำกำไรในปีก่อนหน้าเป็นปัจจัยสำคัญต่อการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวมในปี 2567 โดยเฉพาะอัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม ซึ่งมีความสำคัญสูงสุดในแบบจำลอง Random Forest สะท้อนให้เห็นว่าประสิทธิภาพในการบริหารต้นทุนและค่าใช้จ่ายในการดำเนินงานเป็นสัญญาณสำคัญของผลการดำเนินงานทางการเงินในอนาคต

ผลการทดสอบสมมติฐาน แบ่งออกเป็น 2 ประเด็น ได้แก่ การทดสอบความสัมพันธ์ระหว่างอัตราส่วนทางการเงินกับอัตรากำไรสุทธิต่อรายได้รวม และการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ระหว่าง Random Forest และ Multiple Linear Regression ดังแสดงในตารางที่ 4

ตารางที่ 4 สรุปผลการทดสอบสมมติฐาน

สมมติฐาน	วิธีการทดสอบ	ผลการทดสอบ	การสรุปผล
สมมติฐานที่ 1	Spearman Rank Correlation	พบตัวแปรที่มีความสัมพันธ์อย่างมีนัยสำคัญ 41 ตัวแปร จากทั้งหมด 86 ตัวแปร หรือ ร้อยละ 47.7	สนับสนุน
สมมติฐานที่ 2	เปรียบเทียบค่า RMSE, MAE และ R ² ร่วมกับ Wilcoxon Signed-Rank Test และ Diebold-Mariano Test	RF มีค่า R ² สูงกว่า และมีค่า RMSE, MAE ต่ำกว่า MLR โดย Wilcoxon p = 0.034 แต่ Diebold-Mariano p = 0.232	สนับสนุนบางส่วน

จากตารางที่ 4 ผลการทดสอบสมมติฐานที่ 1 พบว่า อัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจมีความสัมพันธ์กับอัตรากำไรสุทธิต่อรายได้รวมอย่างมีนัยสำคัญทางสถิติ โดยมีตัวแปรจำนวน 41 ตัวแปรจากทั้งหมด 86 ตัวแปร หรือร้อยละ 47.7 ที่มีความสัมพันธ์กับตัวแปรตาม จึงสรุปได้ว่าสมมติฐานที่ 1 ได้รับการสนับสนุน

ผลการทดสอบสมมติฐานที่ 2 พบว่า แบบจำลอง Random Forest มีค่า R² สูงกว่า และมีค่า RMSE รวมถึง MAE ต่ำกว่าแบบจำลอง Multiple Linear Regression แสดงให้เห็นว่า Random Forest มีแนวโน้มให้ผลการพยากรณ์ที่แม่นยำกว่า นอกจากนี้ ผลการทดสอบ Wilcoxon Signed-Rank Test พบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 อย่างไรก็ตาม ผลการทดสอบ Diebold-Mariano Test ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ ดังนั้น จึงสรุปว่าสมมติฐานที่ 2 ได้รับการสนับสนุนบางส่วน

โดยสรุป ผลการทดสอบสมมติฐานสะท้อนว่า ตัวแปรอิสระที่ประกอบด้วยอัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ มีความสัมพันธ์กับผลการดำเนินงานของบริษัท และแบบจำลอง Random Forest มีแนวโน้มเหมาะสมกว่าสำหรับการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวมในบริบทของข้อมูลทางการเงินที่มีความซับซ้อน อย่างไรก็ตาม ผลการเปรียบเทียบแบบจำลองควรตีความด้วยความระมัดระวัง เนื่องจากจำนวนข้อมูลในชุดทดสอบมีขนาดจำกัด

อภิปรายผลการวิจัย

ผลการวิจัยตามกระบวนการที่ 1 ค้นหาแง่มุมด้านค่าใช้จ่ายและผลกำไรจากส่วนประกอบสูงสุดกับอัตรากำไรต่อรายได้รวมสะท้อนว่าต้นทุนและประสิทธิภาพการวิจัยในนี้แสดงการอธิบายรายละเอียดตามแนวคิดของ Fama และ French (2000) ที่ตรวจสอบข้อมูลและวิเคราะห์ผลกำไรในเรื่องนี้ หลักกำไรจากประสบการณ์ปีก่อนที่จะค้นพบกับกำไรสุทธิในปีถัดไปในผลการวิจัยนี้เพิ่มเติมและอีกนัยหนึ่ง Lev และ Thiagarajan (1993) ที่ชี้ว่าแนวคิดด้านต้นทุนการเจริญเติบโตและการพยากรณ์กำไรมากกว่าปัจจัยมหภาคในระยะสั้นๆ ไปในทิศทางเดียวกับผลการวิจัยที่มีความสำคัญต่อสุขภาพของมหภาคและระดับเล็กน้อยกว่าระบบภายในกิจการ

ผลการวิจัยตามวิธีที่ 2 ค้นหา Random Forest ได้อย่างมีประสิทธิภาพในการคาดการณ์ข้อมูลการถดถอยเชิงเส้นพหุคูณ ($R^2 = 0.776$ ต่อ -0.809) คุณสมบัติของข้อมูลเป็นเพียงข้อมูลมิติสูง (86 วิจัยต่อ 40 บริษัท) ไตรฟข้อมูลสามารถจับได้ดีพอๆ กับ Random Forest ที่ใช้ Ensemble Learning ผลการวิจัยของ Breiman (2001) ที่เสนอว่า Random Forest ลดความพิเศษของสัญญาณอย่างละเอียดคือค่าความลาดเคลื่อนของสัญญาณนี้ในการวิจัยด้านล่าง MLR รังสี Gu, Kelly, และ Xiu (2020) ที่ Machine Learning มีแนวทางเหนือรายงานในข้อมูลข่าวสารและผลในการวิจัยนี้แสดงทิศทางเดียวกันในรายงานข่าวไทยและระบบปฏิบัติการ Lim, Hooy, และ Zhu (2023) ที่ การเรียนรู้ทั้งมวลโดยคาดการณ์การพยากรณ์อัตรากำไรสุทธิ เป็นหลักในมุมมองของอุตสาหกรรมและขนาดข้อมูลรวมกับการวิจัยนี้อย่างไรก็ตาม ผลการทดสอบ Diebold-Mariano ยังไม่พบว่าการเปิดตัวอย่างเป็นทางการเนื่องจากขนาดชุดทดสอบที่จำกัด ระบบควรรีความความของ Random Forest รอให้คุณติดตามสนับสนุนอีกครั้ง

ผลการวิจัยตามวิธีที่ 3 โดยทั่วไปมักจะมีความสำคัญที่สุด 3 คือคือค่าใช้จ่ายที่เพิ่มขึ้นของอัตราส่วนกำไรจากโมดูลและอัตรากำไรสุทธิต่อปีที่รวมกันอธิบายคุณลักษณะความสำคัญได้ตามความต้องการ 59.92 สะท้อนว่าผลที่เพิ่มมากขึ้นต่อเนื่องต่อผลการสำรวจผลการวิจัยเกี่ยวกับวิธีการที่แนวคิด Earnings Persistence ของ Ou และ Penman (1989) คุณสมบัติทั่วไปของผลกำไรต่อปีรุ่นที่ในการวิจัยนี้ฟังการพยากรณ์โดยตรงของ Schmedders และ Söderlind (2022) ที่เปิดเผยแง่มุมต่างๆ การทำกำไรเป็นส่วนประกอบหลักในการทำนายผลด้วย Machine Learning เป็นผลในการวิจัยที่ชี้ให้เห็นกลุ่มเดียวกันเป็นวิจารณ์และตรวจ Chen, Cronqvist, Ni, และ Zhang (2022) ที่หาข้อมูลด้านค่าใช้จ่ายค้นคว้าวิจัยวิจัยสูงในแวดวง Machine Learning ที่สำคัญที่สุดคือประวัติความเป็นมาของ Feature Importance ในการวิจัยนี้อีกด้วย อธิบายว่าเหตุใดจึงมีประโยชน์เชิงปฏิบัติในการชี้แนะให้ผู้บริหารและนักลงทุนและติดตามกลุ่มนี้เป็นอันดับแรก

โดยสรุปผลนี้ความเข้มข้นของค่าใช้จ่ายที่เพิ่มขึ้นและผลกำไรเป็นปัจจัยหลักที่สัมพันธ์กับผลเบื้องต้นและ Random Forest เป็นหลักมากกว่า MLR ในการคาดการณ์ข้อมูลจำนวนที่มีความคิดเห็นสูงอย่างไรก็ตามผลการวิจัยควรรีความด้วยสาเหตุเนื่องจากขนาดกลุ่มตัวอย่างและชุดทดสอบที่มีจำนวนจำกัด

องค์ความรู้ใหม่จากการวิจัย

จากการวิจัยครั้งนี้ ทำให้เกิดองค์ความรู้ใหม่ในรูปแบบ “กรอบการพยากรณ์ผลการดำเนินงานด้วยอัตราส่วนทางการเงินและแบบจำลองการเรียนรู้ของเครื่อง” หรือ Financial Ratio–Machine Learning Forecasting Framework ซึ่งแสดงให้เห็นว่า การพยากรณ์อัตรากำไรสุทธิ (Net Profit Margin) ของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มไม่ควรพิจารณาเฉพาะอัตราส่วนทางการเงินแบบแยกส่วน แต่ควรบูรณาการข้อมูลทางการเงินในอดีต ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ เข้ากับแบบจำลองพยากรณ์ที่สามารถรองรับความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้นได้ โดยเฉพาะแบบจำลอง Random Forest ซึ่งไม่เพียง

ช่วยเพิ่มความแม่นยำในการพยากรณ์ แต่ยังสามารถระบุตัวแปรสำคัญที่มีอิทธิพลต่อผลการดำเนินงานผ่านค่า Feature Importance ได้อย่างเป็นระบบ

องค์ความรู้ใหม่จากงานวิจัยนี้สามารถสรุปได้เป็น 3 องค์ประกอบสำคัญ ได้แก่

1. **ปัจจัยนำเข้าในการพยากรณ์** ประกอบด้วยอัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ โดยเฉพาะตัวแปรที่เกี่ยวข้องกับค่าใช้จ่ายดำเนินงาน กำไรจากการดำเนินงาน และกำไรสุทธิในอดีต ซึ่งเป็นกลุ่มตัวแปรที่มีความสัมพันธ์และมีความสำคัญต่อการพยากรณ์อัตรากำไรสุทธิ

2. **กลไกการวิเคราะห์และพยากรณ์** การเปรียบเทียบแบบจำลองพบว่า Random Forest มีความเหมาะสมมากกว่า Multiple Linear Regression ในบริบทของข้อมูลทางการเงินที่มีจำนวนตัวแปรสูงและมีความสัมพันธ์ซับซ้อน เนื่องจากสามารถลดความคลาดเคลื่อนในการพยากรณ์ รองรับความสัมพันธ์ที่ไม่เป็นเชิงเส้น และระบุตัวแปรสำคัญได้อย่างมีประสิทธิภาพ

3. **ผลลัพธ์เชิงการตัดสินใจ** ผลการพยากรณ์สามารถนำไปใช้สนับสนุนการตัดสินใจของนักลงทุน ผู้บริหาร และนักวิเคราะห์ทางการเงิน โดยช่วยชี้ให้เห็นว่าการควบคุมค่าใช้จ่ายดำเนินงานและการรักษาความสามารถในการทำกำไรจากการดำเนินงานเป็นปัจจัยสำคัญต่อผลการดำเนินงานในอนาคต ดังภาพที่ 2



ภาพที่ 2 องค์ความรู้ใหม่จากการวิจัย

จากภาพที่ 2 องค์ความรู้ใหม่ของงานวิจัยนี้อยู่ที่การเชื่อมโยงข้อมูลอัตราส่วนทางการเงินในอดีต ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจ เข้ากับแบบจำลองการพยากรณ์สมัยใหม่ เพื่อใช้คาดการณ์ผลการดำเนินงานของบริษัทในอนาคต ความแตกต่างจากแนวทางการวิเคราะห์ทางการเงินแบบดั้งเดิมคือ งานวิจัยนี้ไม่ได้หยุดอยู่กับการอธิบายผลการดำเนินงานย้อนหลัง แต่พัฒนาไปสู่การใช้ข้อมูลทางการเงินเพื่อพยากรณ์ผลลัพธ์ในอนาคต และระบุตัวแปรที่ควรให้ความสำคัญในการตัดสินใจ

ดังนั้น องค์ความรู้ใหม่จากการวิจัยนี้จึงสามารถนำไปประยุกต์ใช้เป็นแนวทางในการพัฒนาเครื่องมือวิเคราะห์ทางการเงิน การประเมินความสามารถในการทำกำไร และการสนับสนุนการตัดสินใจเชิงกลยุทธ์ของ

บริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม รวมถึงสามารถต่อยอดไปยังกลุ่มอุตสาหกรรมอื่นในตลาดหลักทรัพย์แห่งประเทศไทยได้ในอนาคต

สรุป

ผลการวิจัยพบว่า ตัวแปร 41 จาก 86 ตัวแปรมีความสัมพันธ์อย่างมีนัยสำคัญ โดยอัตราค่าใช้จ่ายดำเนินงานและอัตรากำไรจากการดำเนินงานเป็นปัจจัยสำคัญที่สุด สะท้อนว่าการควบคุมต้นทุนเป็นตัวกำหนดหลักของผลการดำเนินงาน ด้านการพยากรณ์ Random Forest มีประสิทธิภาพสูงกว่า Multiple Linear Regression อย่างชัดเจน (ค่า R^2 สูงกว่า ค่า MAE และ RMSE ต่ำกว่า) และระบุตัวแปรสำคัญสุดคือค่าใช้จ่ายและกำไรจากการดำเนินงานของปีล่าสุด ผลการวิจัยจึงสนับสนุนว่าการบูรณาการอัตราส่วนทางการเงินกับ Machine Learning ช่วยยกระดับการพยากรณ์ผลการดำเนินงานได้ และสามารถใช้เป็นแนวทางสนับสนุนการวิเคราะห์และตัดสินใจลงทุนในอุตสาหกรรมนี้ได้ ทั้งนี้ควรพิจารณาข้อจำกัดด้านขนาดกลุ่มตัวอย่างร่วมด้วย

ข้อเสนอแนะ

จากผลการวิจัย ผู้วิจัยมีข้อเสนอแนะ ดังนี้

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

ผลจากการวิจัยวัตถุประสงค์ที่ 1 พบว่า อัตราส่วนทางการเงิน ปัจจัยภายในของกิจการ และปัจจัยมหภาคทางเศรษฐกิจหลายตัวมีความสัมพันธ์กับอัตรากำไรสุทธิต่อรายได้รวมอย่างมีนัยสำคัญ โดยเฉพาะอัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวมและอัตรากำไรจากการดำเนินงานต่อรายได้รวม ดังนั้น ผู้บริหารบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่มควรให้ความสำคัญกับการควบคุมค่าใช้จ่ายในการดำเนินงาน การบริหารต้นทุน และการเพิ่มประสิทธิภาพของกระบวนการดำเนินงาน เพื่อรักษาระดับความสามารถในการทำกำไรของบริษัทอย่างต่อเนื่อง นอกจากนี้ นักลงทุนและนักวิเคราะห์ทางการเงินควรใช้อัตราร้อยละดังกล่าวเป็นตัวชี้วัดสำคัญในการประเมินแนวโน้มผลการดำเนินงานของบริษัท

ผลจากการวิจัยวัตถุประสงค์ที่ 2 พบว่า แบบจำลอง Random Forest มีประสิทธิภาพในการพยากรณ์อัตรากำไรสุทธิต่อรายได้รวมสูงกว่าแบบจำลอง Multiple Linear Regression ดังนั้น บริษัท นักลงทุน และนักวิเคราะห์ทางการเงินควรพิจารณานำแบบจำลองการเรียนรู้ของเครื่องมาประกอบกับการวิเคราะห์ทางการเงิน โดยเฉพาะในกรณีที่ข้อมูลมีจำนวนตัวแปรมากหรือมีความสัมพันธ์ซับซ้อน อย่างไรก็ตาม การนำแบบจำลองดังกล่าวไปใช้ควรพิจารณาพร้อมกับความรู้เชิงธุรกิจ การตีความผลลัพธ์ทางการเงิน และข้อจำกัดของข้อมูล เพื่อให้การตัดสินใจมีความรอบคอบและเหมาะสมกับบริบทของแต่ละบริษัท

ผลจากการวิจัยวัตถุประสงค์ที่ 3 พบว่า ตัวแปรที่มีความสำคัญต่อการพยากรณ์ผลการดำเนินงาน ได้แก่ อัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม อัตรากำไรจากการดำเนินงานต่อรายได้รวม และอัตรากำไรสุทธิต่อรายได้รวมในอดีต ดังนั้น ผู้บริหารควรใช้ตัวแปรเหล่านี้เป็นสัญญาณเตือนทางการเงินและเป็นข้อมูลประกอบการกำหนดกลยุทธ์ เช่น การควบคุมค่าใช้จ่าย การปรับปรุงประสิทธิภาพการผลิต การบริหารต้นทุนวัตถุดิบ และการรักษาความสามารถในการทำกำไรจากการดำเนินงาน เพื่อส่งเสริมผลการดำเนินงานในระยะยาว

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

งานวิจัยนี้ได้ข้อค้นพบที่สำคัญ คือ อัตราค่าใช้จ่ายดำเนินงานต่อรายได้รวม (M) และอัตรากำไรจากการดำเนินงานต่อรายได้รวม (G) เป็นตัวแปรที่มีอิทธิพลสูงสุดต่อการพยากรณ์อัตรากำไรสุทธิของบริษัทในกลุ่มอุตสาหกรรมอาหารและเครื่องดื่ม และแบบจำลอง Random Forest มีความสามารถในการพยากรณ์ที่เหนือกว่าแบบจำลองการถดถอยพหุคูณ (MLR) อย่างชัดเจนในบริบทข้อมูลที่มีจำนวนตัวแปรมากกว่าจำนวนกลุ่มตัวอย่าง

สามารถนำไปประยุกต์ใช้กับการวางแผนบริหารต้นทุนดำเนินงานและการประเมินความสามารถในการทำกำไรของบริษัทในอุตสาหกรรมอาหารและเครื่องดื่ม โดยควรให้ความสำคัญกับการควบคุมประสิทธิภาพค่าใช้จ่ายดำเนินงานและการรักษาอัตรากำไรจากการดำเนินงานให้มีความต่อเนื่อง สำหรับประเด็นในการวิจัยครั้งต่อไปควรทำวิจัยในประเด็นเกี่ยวกับการขยายกลุ่มตัวอย่างไปยังหลายอุตสาหกรรม การเพิ่มตัวแปรด้านเศรษฐกิจมหภาค และ ESG และการทดสอบแบบจำลองทางเลือกอื่น เช่น XGBoost หรือ Neural Network ร่วมกับการตรวจสอบความแม่นยำแบบ Walk-Forward Validation

เอกสารอ้างอิง

- ธนวัฒน์ ศรีสมบูรณ์ และ วรณา จันทรเพ็ญ. (2566). การเปรียบเทียบประสิทธิภาพแบบจำลอง Machine Learning แบบ Tree-based Ensemble ในการพยากรณ์ผลการดำเนินงานของบริษัทจดทะเบียนในตลาดหลักทรัพย์. *วารสารเศรษฐศาสตร์และบริหารธุรกิจ*, 15(2), 45–63.
- นริศรา ประไพ, ธนกฤต วงศ์ศิริภากร และ พิมพ์ชนก เนตรศิริ. (2565). การเปรียบเทียบประสิทธิภาพของแบบจำลอง Random Forest และ Multiple Linear Regression ในการพยากรณ์กำไรสุทธิของบริษัทในตลาดหลักทรัพย์แห่งประเทศไทย. *วารสารการเงิน การบัญชี และการจัดการ*, 13(2), 78–95.
- วิชัย อนันต์โกศล. (2567). การเปรียบเทียบประสิทธิภาพแบบจำลองทางสถิติและ Machine Learning ในการพยากรณ์กำไรสุทธิของบริษัทในกลุ่มอุตสาหกรรมเกษตรและอาหาร ตลาดหลักทรัพย์แห่งประเทศไทย. *วารสารบริหารธุรกิจและสังคมศาสตร์ มหาวิทยาลัยรามคำแหง*, 7(1), 85–102.
- วิศรุต แก้วมหา และวิศ ปญญาฉัตรพร. (2564). การคาดการณ์ผลตอบแทนในอนาคตของตราสารทุนหุ้นสามัญโดยใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง. *วารสารนวัตกรรมการธุรกิจ การจัดการ และสังคมศาสตร์*, 2(3), 108–123.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227.
- Cao, J., Chen, X., & Wang, Y. (2020). Cross-validated machine learning models for earnings forecasting in emerging markets. *Journal of Emerging Market Finance*, 19(3), 301–325.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15(1), 3133–3181.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.
- Hezam, Y., Luong, H., & Anthonysamy, L. (2025). Machine learning in predicting firm performance: A systematic review. *China accounting and finance review*, 27(3), 309–339.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques: A review. *European Journal of Operational Research*, 180(1), 1–28.
- Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5), 1541–1578.

- Nissim, D., & Penman, S. H. (2001). Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies*, 6(1), 109–154.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.
- Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York: Free Press.
- Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 71(3), 289–315.
- Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review*, 83(3), 823–853.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
- Tutcu, B., Kayakuş, M., Terzioğlu, M., Ünal Uyar, G. F., Talaş, H., & Yetiz, F. (2024). Predicting financial performance in the IT industry with machine learning: ROA and ROE analysis. *Applied Sciences*, 14(17), 7459.
- Wasserbacher, H., & Spindler, M. (2022). Machine learning for financial forecasting, planning and analysis: Recent developments and pitfalls. *Digital Finance*, 4(1), 63–88.