ตัวแบบการพยากรณ์ปริมาณการจัดเก็บวัสดุอะไหล่ในคลังโดยใช้เทคนิคการเรียนรู้ของเครื่อง: กรณีศึกษาบริษัทผลิตปูนซีเมนต์

FORECASTING MODEL FOR THE AMOUNT OF SPARE PARTS STORAGE IN THE WAREHOUSE USING MACHINE LEARNING TECHNIQUES: CASE STUDY OF A CEMENT PLANT

ดวงเดือน อ่องรุ่งเรือง ศุภศิริ ไทยวิรัช และวันทนี ประจวบศุภกิจ*

Duongduen Ongrungrueng Suphasiri Thaiwirach and Wanthanee Prachuabsupakij*

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

Department of Information Technology, Faculty of Industrial Technology and Management,

King Mongkut's University of Technology North Bangkok Prachinburi Campus

*corresponding author e-mail: wanthanee.p@fitm.kmutnb.ac.th

(Received: 5 March 2020; Revised: 18 April 2020; Accepted: 19 April 2020)

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบในการพยากรณ์ปริมาณการจัดเก็บวัสดุคงเหลือแต่ละวัสดุตามปริมาณที่เหมาะสมด้วยการพยากรณ์หาจำนวนต่ำสุดและสูงสุดในการจัดการคลังวัสดุโดยใช้เทคนิคการเรียนรู้ของเครื่อง ข้อมูลที่นำมาใช้เป็นข้อมูลวัสดุอะไหล่ที่ใช้งานจริงของบริษัทผลิตปูนซีเมนต์แห่งหนึ่งในประเทศไทย จำนวนทั้งสิ้น 7,258 ระเบียน โดยผู้วิจัยเลือกอัลกอริทึมในการพยากรณ์ที่ใช้ในการวิจัยครั้งนี้ จำนวน 5 อัลกอริทึม ได้แก่ แรนดอมฟอเรสต์ มัลติเลเยอร์เพอร์เซฟตรอน เคเนียร์เรสเนเบอร์ เคสตาร์ และแรนดอมทรี จากผลการทดลองการเปรียบเทียบประสิทธิภาพอัลกอริทึมที่ใช้ในการพยากรณ์ปริมาณการจัดเก็บวัสดุในคลัง พบว่าอัลกอริทึมที่ให้ค่าประสิทธิภาพที่ดีที่สุดในการพยากรณ์ค่าสูงสุด คือ อัลกอริทึมแรนดอมฟอเรสต์ โดยมีค่ารากที่สองของความคลาดเคลื่อนที่ดีที่สุด เท่ากับ 2.040 ในขณะที่อัลกอริทึมแรนดอมทรีให้ค่ารากที่สองของความคลาดเคลื่อนที่ใกล้เคียงกัน คือ 2.117 สำหรับการพยากรณ์ค่าสูงสุด ในขณะที่ผลการพยากรณ์ค่าต่ำสุด พบว่า อัลกอริทึมแรนดอมทรีมีค่ารากที่สองของความคลาดเคลื่อนที่ดีที่สุดเท่ากับ 0.803 ส่วนอัลกอริทึมแรนดอมฟอเรสต์มีค่ารากที่สองของความคลาดเคลื่อนเท่ากับ 0.906 อย่างไรก็ตามในงานวิจัยนี้ได้เลือกใช้อัลกอริทึมแรนดอมฟอเรสต์ ในการพยากรณ์ทั้งค่าสูงสุดและต่ำสุด เนื่องจากให้ค่าประสิทธิภาพที่เหมาะสมที่สุดเมื่อพิจารณาทั้งผลการพยากรณ์ข้อมูลใหม่และค่ารากที่สองของความคลาดเคลื่อน

คำสำคัญ: การสร้างตัวแบบ การพยากรณ์ การเรียนรู้ของเครื่อง แรนดอมฟอเรสต์ การจัดการคลังวัสดุ

Abstract

The research aims to construct the forecasting model for the remaining amount of spare parts storage at suitable inventory level to forecast minimum and maximum levels for material inventory management using machine learning techniques. This research collected from a real data of a cement factory in Thailand. The dataset has 7,258 instances. We selected five algorithms for forecasting the data including Random forest, Multi–layer perceptron, K–nearest neighbor, K–star, and Random tree. The results of the comparison performance of algorithms for forecasting the amount of spare parts storage in the warehouse showed that the Random forest algorithm provided the highest root mean square error (RMSE) at 2.040 while the Random tree algorithm provided the similar RMSE at 2.117 for forecasting MAX value (Model MAX). On the other hand, for forecasting MIN value, Random tree algorithm obtained the highest RMSE at 0.803 while the Random forest algorithm showed the RMSE at 0.906. However, in this paper, the Random forest algorithm is selected as a forecasting model of both MIN and MAX values because it has the most appropriate performance value both the results of forecasting on unseen data and RMSE values.

Keywords: Model, Forecasting, Machine learning, Random forest, Material inventory management

บทนำ

ในปัจจุบันการจัดการคลังพัสดุมีความสำคัญต่อการทำธุรกิจของทุกบริษัท ซึ่งต้องมีการวิเคราะห์แนวทางที่เหมาะสมที่สุดในการควบคุมคลังพัสดุให้มีปริมาณสินค้าที่เหมาะสมจึงจะทำให้ธุรกิจดำเนินได้อย่างต่อเนื่องและไม่ก่อให้เกิดความสูญเสียตามมา เช่นเดียวกับในกรณีศึกษาซึ่งเป็นบริษัทผลิตปูนซีเมนต์รายหนึ่งของประเทศไทยที่มีหน่วยงานจัดหาและพัสดุเป็นหน่วยงานที่ดูแลควบคุมและประสานงานบริหารงานจัดหาและบริหารงานพัสดุ รวมทั้งจัดเก็บวัสดุทั่วไป วัสดุในกระบวนการผลิต และวัสดุอะไหล่ ให้พร้อมและเพียงพอเพื่อใช้ในกระบวนการผลิต และการซ่อมบำรุงเครื่องจักรต่างๆ ในกระบวนการผลิตเพื่อที่จะไม่กระทบกับการผลิตผลิตภัณฑ์ แต่จากมูลค่าและปริมาณในการจัดเก็บวัสดุอะไหล่ (Spare parts) ได้แก่ อะไหล่ไฟฟ้า

อะไหล่เครื่องกล และอะไหล่เครื่องจักรกลเหมือง มีมูลค่าสูงเกินความจำเป็น อีกทั้งอะไหล่บางรายการมีปริมาณการจัดเก็บไว้เป็นปริมาณมากและไม่สอดคล้องกับการเบิกนำไปใช้งาน จึงทำให้วัสดุอะไหล่เหลือในคงคลังเป็นจำนวนมาก ก่อให้เกิดมูลค่าการเก็บรักษาที่ไม่คุ้มค่า และมีต้นทุนในการเก็บวัสดุคงคลัง ดังนั้นทางโรงงานจึงพิจารณาที่จะจัดทำระบบในการควบคุมและติดตามโดยการวิเคราะห์ข้อมูลการใช้และการจัดเก็บวัสดุอะไหล่เหล่านี้ให้มีความเหมาะสมสอดคล้องกับเป้าหมายที่กำหนดเพื่อตอบสนองความต้องการใช้วัสดุอะไหล่ภายใต้เงื่อนไขที่มีความแน่นอนและเงื่อนไขที่จำเป็นต้องจัดเก็บในปริมาณที่น้อยลง โดยมีการทบทวนระดับปริมาณการจัดเก็บต่ำสุดและสูงสุด (MIN/MAX) ของวัสดุอะไหล่ที่จำเป็นต้องจัดเก็บอย่างเหมาะสมเพื่อให้มีระดับการให้บริการ (Service level) ตามที่ผู้ใช้งานต้องการได้อย่างมีประสิทธิภาพสูงสุด

จากปัญหาสินค้าคงคลังข้างต้นพบว่ามีงานวิจัยจำนวนมากได้นำเสนอเทคนิคต่างๆ มาช่วยแก้ไขปัญหา เช่น งานวิจัยของ Sawarni (2018) ได้นำเสนอวิธีการจัดการสินค้าคงคลังของวัสดุอะไหล่ของบริษัทแห่งหนึ่งในอินโดนีเซียด้วยหลักการบริหารจัดการโดยใช้สูตรคำนวณ ซึ่งบริษัทต้องการลดค่าใช้จ่ายในการจัดเก็บสินค้าในคลังโดยผู้วิจัยได้ใช้เทคนิค ABC analysis ในการจำแนกประเภทของวัสดุแต่ละรายการ หลังจากนั้นคำนวณจุดต่ำสุด ซึ่งผลจากการคำนวณ พบว่า ถ้าอัตราการหมุนเวียนของวัสดุคงเหลือเพิ่มขึ้น และค่าเฉลี่ยของปริมาณวัสดุและต้นทุนรวมของการจัดเก็บวัสดุลดลงจะทำให้ลดต้นทุนในการจัดเก็บวัสดุลงได้เมื่อเทียบกับวิธีการจัดการสินค้าคงคลังแบบอื่นๆ ที่นำมาเปรียบเทียบ นอกจากนี้ยังพบว่า การใช้เทคนิคการเรียนรู้ของเครื่องซึ่งเป็นการให้เครื่องคอมพิวเตอร์หรือเครื่องจักรเรียนรู้งานใดงานหนึ่งจากตัวอย่างหรือประสบการณ์จำนวนหนึ่งเพื่อให้ทำงานนั้นๆ ได้อย่างมีประสิทธิภาพ (ฤทธิไกร, 2559) ได้ถูกนำมาประยุกต์ใช้ในงานด้านต่างๆ เช่น ด้านการศึกษา ด้านอุตสาหกรรม และด้านธุรกิจ สำหรับอัลกอริทึมที่นิยมนำมาประยุกต์ใช้กับงานด้านต่างๆ มีหลายอัลกอริทึม ตัวอย่างเช่น แรนดอมฟอเรสต์ (Random forest) จะสร้างตัวแบบจากต้นไม้ตัดสินใจหลายๆ ตัวแบบย่อยๆ โดยแต่ละตัวแบบจะได้รับชุดข้อมูลที่ไม่เหมือนกัน (Random sampling) โดยข้อมูลย่อยจะถูกแบ่งออกมาเป็น n ชุด ที่ไม่เหมือนกันตามจำนวนต้นไม้ตัดสินใจ และจะสร้างตัวแบบเพื่อทำการพยากรณ์คำตอบของตนเองด้วยการโหวต อัลกอริทึมมัลติเลเยอร์เพอร์เซฟตรอน (Multi-layer perceptron) จะใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) สำหรับการฝึกฝนกระบวนการส่งค่าย้อนกลับ (บทเรียนออนไลน์ โครงข่ายประสาทเทียม, 2563) เทคนิคแรนดอมทรี (Random tree) ซึ่งเป็นเทคนิคที่ใช้ในการจำแนกหมวดหมู่เช่นเดียวกับ C4.5 โดยมีหลักการสร้าง Tree จากการสุ่ม Tree หลายๆ แบบ ในแต่ละโหนดแล้วเลือกมาประมวลผลโดยไม่ใช้การ Prune (พัฒนพงษ์, 2560) วิธีเคเนียร์เรสเนเบอร์ (K-nearest neighbor) จะจำแนกโดยดูว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน (“K” ตัว) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือ

ใกล้เคียงกันมากที่สุด โดยจะหาผลรวมของจำนวนเงื่อนไขสำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด การนำเทคนิคของอัลกอริทึมเพื่อนบ้านใกล้ที่สุดไปใช้นั้น เป็นการหาระยะห่างระหว่างแต่ละตัวแปร (Attribute) ในข้อมูล แล้วหาผลรวมของระยะห่างทั้งหมด แล้วพิจารณาเลือกคำตอบจากจำนวน K ตัว (บทเรียนออนไลน์อัลกอริทึมการค้นหาเพื่อนบ้านใกล้สุด k ตัว, 2562) และเคสตาร์ (K-star) ซึ่งเป็นอัลกอริทึมที่ใช้การเรียนรู้ที่อาศัยชุดข้อมูลเป็นฐานชนิดหนึ่ง โดยจะใช้การวัดระยะห่างระหว่างข้อมูลจุดต่างๆ แต่แตกต่างจากวิธีการอื่นๆ คือ ใช้ค่าเอนโทรปีเป็นตัววัดระยะห่าง และมีการคำนวณค่าความน่าจะเป็นระหว่างข้อมูลใหม่กับข้อมูลที่มีอยู่ทั้งหมดเพื่อกำหนดประเภทให้กับข้อมูลใหม่ โดยเลือกประเภทที่มีค่าความน่าจะเป็นสูงสุด (Yoan et al., 2016) เป็นต้น

โดยเพาะอย่างยิ่งการนำเทคนิคการเรียนรู้ของเครื่องไปประยุกต์ใช้กับภาคอุตสาหกรรม อาทิ การพยากรณ์กระแสไฟฟ้าขัดข้อง กรณีศึกษา: การไฟฟ้าส่วนภูมิภาค เขต 3 ภาคกลาง โดยใช้วิธี K-mean และการจำแนกประเภทด้วยวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อให้ตัวแบบพยากรณ์มีประสิทธิภาพสูงสุด (ทิตา, 2559) รวมทั้งงานวิจัยของ อรรณพ (2562) ที่ได้นำเสนอการพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซลด้วยเทคนิคการเรียนรู้โดยเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซลด้วยตัวแบบโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น การถดถอยเชิงเส้น และซัพพอร์ตเวกเตอร์แมชชีน โดยตัวแบบโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นให้ค่าประสิทธิภาพโดยรวมสูงที่สุด มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 2.38% และการจัดการสินค้าคงคลังเพื่อพยากรณ์วัสดุคงค้างในคลังสินค้าของ Radrigo et al. (2017) ซึ่งได้ใช้ข้อมูลจาก Kaggle's competition มาเป็นข้อมูลที่ใช้ในการแข่งขันเทคนิคการเรียนรู้ของเครื่องจักร โดยมีข้อมูลทั้งสิ้น 1,929,936 ระเบียน โดยได้ทำการทดลองและเปรียบเทียบประสิทธิภาพกับอัลกอริทึมที่หลากหลาย ประกอบด้วย Logistic regression CART Random forest Gradient tree boosting และ Bagging ผลการทดลอง พบว่า Gradient tree boosting เป็นอัลกอริทึมที่ให้ค่าความถูกต้องสูงสุด แต่ในทางตรงกันข้ามเมื่อพิจารณาที่ค่าความเที่ยงตรงและค่าระลึกพบว่า Bagging ให้ค่าที่ดีที่สุด เป็นต้น

จากงานวิจัยที่เกี่ยวข้องข้างต้นพบว่าเทคนิคการเรียนรู้ของเครื่องได้ถูกนำมาประยุกต์ใช้กับงานจัดการสินค้าคงคลัง แต่ยังไม่มีงานวิจัยที่นำเทคนิคดังกล่าวมาใช้ในการพยากรณ์ปริมาณการจัดเก็บวัสดุในคลังส่วนของการหาจำนวนสูงสุดและต่ำสุดที่เหมาะสม ดังนั้นผู้วิจัยจึงได้ทำการศึกษาข้อมูลจากบริษัทกรณีศึกษา โดยการนำข้อมูลวัสดุอะไหล่ที่จัดเก็บในโรงงานตั้งแต่ปี พ.ศ. 2558–2562 จำนวนทั้งสิ้น 7,258 ระเบียน และ 113 คุณลักษณะ มาเป็นปัจจัยในการวิเคราะห์เพื่อใช้ในการพยากรณ์ปริมาณการจัดเก็บวัสดุคงเหลือแต่ละวัสดุตามปริมาณที่เหมาะสมด้วย

การพยากรณ์หาจำนวนต่ำสุดและสูงสุดในการจัดเก็บวัสดุในคลังด้วยเทคนิคการเรียนรู้ของเครื่อง เพื่อให้หน่วยงานจัดหาและพัสดุสามารถนำมาใช้ในการควบคุมปริมาณการจัดเก็บวัสดุอะไหล่ที่เหมาะสม และลดค่าใช้จ่ายในการจัดเก็บวัสดุอะไหล่ได้ในอนาคต

วิธีดำเนินการวิจัย

ข้อมูลที่นำมาใช้ในการสร้างตัวแบบในการพยากรณ์เป็นข้อมูลวัสดุอะไหล่ของบริษัทผลิตปูนซีเมนต์แห่งหนึ่ง ในประเทศไทย โดยนำข้อมูลตั้งแต่ปี พ.ศ. 2558–2562 จำนวนทั้งสิ้น 7,258 ระเบียน และ 113 คุณลักษณะ (คอลัมน์) โดยแบ่งเป็นชุดข้อมูลสำหรับขั้นตอนการสร้างตัวแบบพยากรณ์ (ฝึกและทดสอบ) ร้อยละ 70 มีจำนวน 5,181 ระเบียน และชุดข้อมูลใหม่เพื่อทดสอบตัวแบบร้อยละ 30 มีจำนวนข้อมูล 2,077 ระเบียน

1. การเลือกข้อมูล

1.1 ชุดข้อมูลเดิมก่อนทำการเลือกข้อมูล

ทำการเลือก (Select) ข้อมูลจากคลังข้อมูล (Data warehouse) ของบริษัท ซึ่งมี 113 คอลัมน์ โดยได้คัดเลือกเฉพาะปัจจัยที่เป็นตัวเลขมาพิจารณาเพื่อทำการพยากรณ์ข้อมูลค่าสูงสุดและต่ำสุด ซึ่งเหลือข้อมูลเพียง 72 คอลัมน์ ดังตารางที่ 1

ตารางที่ 1 รายละเอียดปัจจัยเดิมทั้งหมด

ที่	ปัจจัย	รายละเอียดปัจจัย
1	On hand quantity	จำนวนของวัสดุที่มีอยู่ในคลังปัจจุบัน
2	On hand amount	มูลค่าของวัสดุที่มีอยู่ในคลังปัจจุบัน
3	Moving avg price	ราคาต่อหน่วยของวัสดุแต่ละชิ้น
4	Standard price	ราคามาตรฐาน
5	On order	จำนวนของวัสดุที่ทำการจัดซื้อ
6	Planned delivery time	จำนวนวันที่ทำการจัดส่งวัสดุที่ได้ทำการจัดซื้อ
7	Reorder point	จำนวนวัสดุที่ควรจัดเก็บขั้นต่ำ
8	Max stock level	จำนวนวัสดุที่ควรจัดเก็บสูงสุด
9	safety stock	ปริมาณวัสดุคงคลังที่มีไว้เผื่อกรณีของขาด
10	Minimum lot size	จำนวนการสั่งซื้อขั้นต่ำ
11	Maximum lot size	จำนวนการสั่งซื้อสูงสุด
12–72	Usage in month	จำนวนการเบิกใช้วัสดุในแต่ละเดือน ตั้งแต่เดือนที่ 1 ถึงเดือนที่ 60 (จำนวน 60 คอลัมน์)

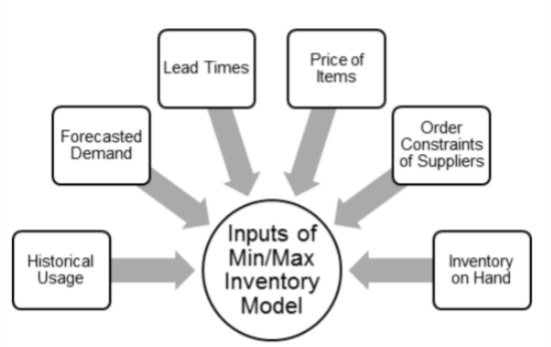
1.2 ชุดข้อมูลที่ถูกคัดเลือก

พิจารณาปัจจัยเบื้องต้นในการจัดเก็บค่าสูงสุดและต่ำสุดโดยการสอบถามผู้ใช้ ซึ่งปรากฏปัจจัยที่เกี่ยวข้องในมุมมองของผู้ใช้ จำนวน 4 ปัจจัย ดังตารางที่ 2

ตารางที่ 2 รายละเอียดปัจจัยที่ผู้ใช้เลือก

ปัจจัย	รายละเอียดปัจจัย
Planned delivery time	จำนวนวันที่ทำการจัดส่งวัสดุที่ได้ทำการจัดซื้อ
Reorder point	จำนวนวัสดุที่ควรจัดเก็บขั้นต่ำ
Max stock level	จำนวนวัสดุที่ควรจัดเก็บสูงสุด
Usage in month	จำนวนการเบิกใช้วัสดุในแต่ละเดือน ตั้งแต่เดือนที่ 1 ถึงเดือนที่ 60

จากตารางที่ 2 พบว่า ปัจจัยที่ผู้ใช้เลือกนั้นมีจำนวนน้อยเกินไปจึงทำให้การนำข้อมูลไปฝึกเพื่อสร้างตัวแบบนั้นมีประสิทธิภาพไม่ดีเท่าที่ควร ดังนั้นผู้วิจัยจึงได้ศึกษาปัจจัยเพิ่มเติมจากงานวิจัยของ Harshard & Priyanka (2015) โดยปัจจัยที่ Harshard ใช้ ดังภาพที่ 1 ดังนั้นผู้วิจัยจึงได้ทำการทดลองเพิ่มปัจจัยในการพยากรณ์ ดังตารางที่ 3



ภาพที่ 1 ตัวอย่างข้อมูลที่นำมาใช้กับโมเดล MIN และ MAX

2. การเตรียมข้อมูล

นำข้อมูลที่รวบรวมมาได้ซึ่งมีทั้งข้อมูลที่ยังเปิดสต็อกไว้และถูกปิดสต็อกไปแล้ว โดยตัวที่ถูกปิดไปแล้วนั้นไม่ได้มีการนำข้อมูลออกจากฐานข้อมูล แต่เป็นการเก็บสถานะแทน โดยจะมีชื่อสถานะว่า Flagged for deletion ซึ่งถ้าในช่องสถานะไหนที่เป็นค่า ‘X’ จะไม่ถูกนำมาใช้เพราะถือว่าเป็นตัวที่ปิดสต็อกไปแล้ว ข้อมูลเหล่านี้จึงถูกลบออกไปจากชุดข้อมูลที่จะนำเข้ามาใช้ในการฝึก ในรายการที่มีข้อมูลรายการวัสดุที่ซ้ำกัน แต่มีค่า Moving avg price ที่แตกต่างกัน จะต้องคัดเลือกว่าควรเลือกใช้ค่า Moving avg price ค่าใด โดยวิธีการคัดเลือกก็จะดูจากสถานะ Valuation type

ว่าถ้ามีสถานะเป็น 'NEW' ก็จะนำข้อมูลตัวนี้มาใช้ แต่ถ้าบางตัวที่ไม่มีสถานะที่เป็น 'NEW' เลยให้เลือกใช้ตัวที่เป็นค่าว่างแทน โดยใช้เทคนิค Generalization เพื่อลดความเฉพาะเจาะจงของข้อมูล โดยการแปลงข้อมูลการเบิกใช้วัสดุ (Usage in month) ทั้งหมด 60 เดือน ให้เป็นรายปี โดยแบ่งเป็น 5 ปี เดือนที่ 1–12 คือ ปีล่าสุดหรือปีที่ 5 เดือนที่ 13–24 คือ ปีที่ 4 เดือนที่ 25–36 คือ ปีที่ 3 เดือนที่ 37–48 คือ ปีที่ 2 และเดือนที่ 49–60 คือ ปีที่ 1 โดยแต่ละปีจะเก็บเป็นค่าเฉลี่ยการเบิกใช้ของปีนั้นๆ ดังนั้นข้อมูลนี้จะถูกลดลงจาก 60 คอลัมน์ เป็น 5 คอลัมน์ เท่านั้น และทำการ Normalize ค่าของ Moving avg price ให้อยู่ในช่วง 0–1 เนื่องจากข้อมูลในคอลัมน์นี้มีช่วงของข้อมูลที่ห่างกันมาก ซึ่งอาจส่งผลกระทบต่อประสิทธิภาพของตัวแบบที่ใช้ในการพยากรณ์

3. การสร้างตัวแบบพยากรณ์

หลังจากผ่านขั้นตอนการเตรียมข้อมูลซึ่งทำให้ได้ข้อมูลที่สมบูรณ์ที่พร้อมสำหรับการนำมาฝึกตัวแบบ จำนวนทั้งสิ้น 5,181 ระเบียน กับ 7 คอลัมน์ ซึ่งข้อมูลนี้จะถูกมาใช้ในการสร้างตัวแบบ 2 ตัว คือ ตัวแบบสำหรับการพยากรณ์จำนวนสูงสุด และตัวแบบสำหรับการพยากรณ์จำนวนต่ำสุด ตัวอย่างปัจจัยสำหรับการสร้างตัวแบบเพื่อพยากรณ์จำนวนสูงสุด ประกอบด้วย 7 ปัจจัย ดังตารางที่ 3

ตารางที่ 3 รายละเอียดปัจจัยสำหรับการสร้างตัวแบบเพื่อพยากรณ์จำนวนสูงสุด

ปัจจัย	รายละเอียดปัจจัย
On hand quantity	จำนวนของวัสดุที่มีอยู่ในคลังปัจจุบัน
Moving avg price	ราคาต่อหน่วยของวัสดุแต่ละชิ้น
On order	จำนวนของวัสดุที่ทำการจัดซื้อ
Planned delivery time	จำนวนวันที่ทำการจัดส่งวัสดุที่ได้ทำการจัดซื้อ
Reorder point	จำนวนวัสดุที่ควรจัดเก็บขั้นต่ำ
Usage in year	ข้อมูลการเบิกใช้วัสดุเฉลี่ย 1 ปี
Max stock level	จำนวนวัสดุที่ควรจัดเก็บสูงสุด (คำเฉลย)

ส่วนปัจจัยสำหรับการสร้างตัวแบบเพื่อพยากรณ์จำนวนต่ำสุด ประกอบด้วย 7 ปัจจัยเช่นกัน ซึ่งรายละเอียดของปัจจัยจะเหมือนกันกับการพยากรณ์จำนวนสูงสุด แต่ต่างกันที่คำเฉลยซึ่งจะเป็นจำนวนที่กำหนดให้จัดเก็บวัสดุต่ำสุด (Min stock level) ซึ่งงานวิจัยนี้ได้เลือกอัลกอริทึมสำหรับการพยากรณ์ปริมาณการจัดเก็บวัสดุคงเหลือ จำนวน 5 อัลกอริทึม ได้แก่ Random forest Multi-layer perceptron K-nearest neighbor K-star และ Random tree

4. การออกแบบการทดสอบและการวัดประสิทธิภาพ

ทำการทดสอบโดยใช้การทดสอบแบบไขว้แบบ 10 ส่วน (10–fold cross validation) และใช้โปรแกรม Weka 3.8.3 เป็นเครื่องมือในการสร้างและวัดประสิทธิภาพตัวแบบ เพื่อประเมินประสิทธิภาพของตัวแบบพยากรณ์ และพิจารณาค่ารากที่สองของความคลาดเคลื่อน (Root mean square error: RMSE) เพื่อให้ได้แบบจำลองที่ดีที่จะนำไปใช้ในการพยากรณ์ สูตรในการคำนวณค่า RMSE ดังสมการที่ 1 (วรางคณา, 2016)

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2} \quad (1)$$

โดยที่ Y_t^s คือ ค่าประมาณจากตัวแบบ

Y_t^a คือ ค่าที่แท้จริง

T คือ จำนวนเวลาที่ใช้ในการประมาณตัวแบบ

ผลการวิจัยและอภิปรายผล

ผลการวิจัย ประกอบด้วย 2 ส่วน ได้แก่ ผลการศึกษาประสิทธิภาพของตัวแบบจากชุดข้อมูลฝึก และผลการศึกษาประสิทธิภาพของตัวแบบจากชุดข้อมูลใหม่

1. ประสิทธิภาพของตัวแบบจากชุดข้อมูลฝึก

จากการทดลองสร้างตัวแบบเพื่อพยากรณ์ปริมาณการจัดเก็บวัสดุคงคลัง โดยแบ่งเป็นตัวแบบสำหรับค่าสูงสุดและค่าต่ำสุด ซึ่งสามารถแสดงค่าประสิทธิภาพของแต่ละอัลกอริทึมสำหรับแต่ละตัวแบบ ดังตารางที่ 4

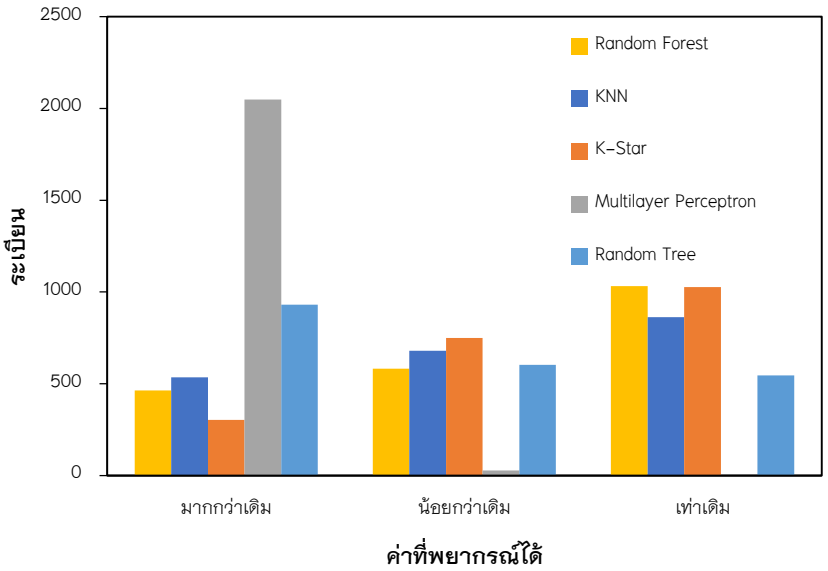
ตารางที่ 4 การเปรียบเทียบประสิทธิภาพด้วยค่า RMSE ของแต่ละอัลกอริทึมเพื่อหาตัวแบบสำหรับพยากรณ์ค่าสูงสุดและค่าต่ำสุด

อัลกอริทึม	ค่าต่ำสุด (MIN)	ค่าสูงสุด (MAX)
Multi–layer Perceptron	5.516	13.962
K–star	1.363	3.423
K–nearest Neighbor	2.136	2.531
Random Tree	0.803	2.117
Random Forest	0.906	2.040

จากตารางที่ 4 พบว่า อัลกอริทึม Random tree เป็นอัลกอริทึมที่ให้ค่าประสิทธิภาพของ RMSE ที่ดีที่สุดในการพยากรณ์จำนวนวัสดุในคลังที่ต่ำที่สุด โดยมีค่า RMSE เท่ากับ 0.803 ในส่วนของอัลกอริทึมที่ให้ค่าประสิทธิภาพ RMSE ที่ดีที่สุดในการพยากรณ์จำนวนวัสดุสูงสุดของวัสดุในคลัง คือ อัลกอริทึม Random forest ซึ่งให้ค่าประสิทธิภาพในการพยากรณ์ เท่ากับ 2.040 ในทางตรงกันข้ามสำหรับชุดข้อมูลชุดนี้ อัลกอริทึมที่ให้ค่าประสิทธิภาพที่ต่ำที่สุด คือ อัลกอริทึม Multilayer perceptron ซึ่งให้ค่า RMSE ของการหาจำนวนต่ำสุดและสูงสุดของวัสดุในคลัง เท่ากับ 5.516 และ 13.962 ตามลำดับ

2. ประสิทธิภาพของตัวแบบจากชุดข้อมูลใหม่

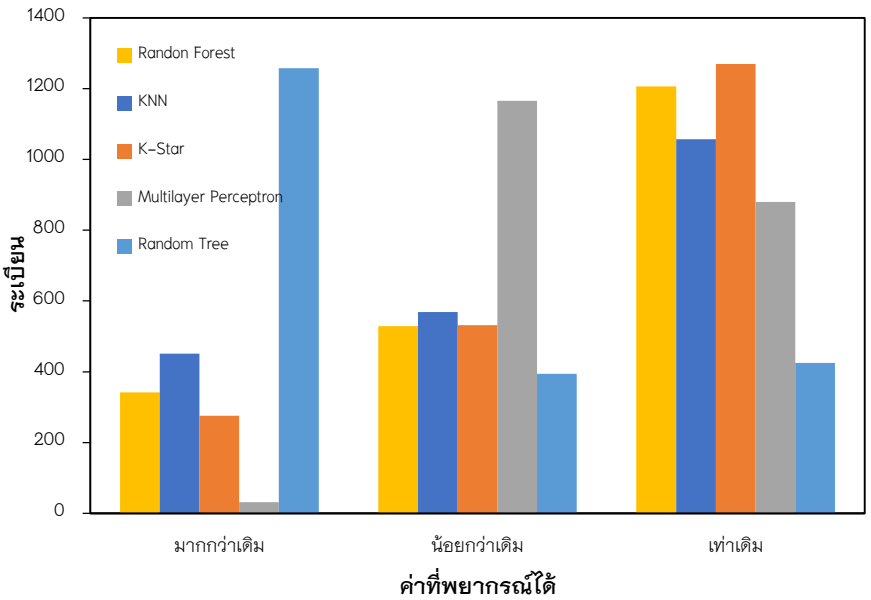
การเปรียบเทียบประสิทธิภาพของตัวแบบกับชุดข้อมูลโดยนำชุดข้อมูลที่ได้ทำการแบ่งไว้ ร้อยละ 30 (2,077 ระเบียน) มาใช้เพื่อทดสอบประสิทธิภาพในการพยากรณ์ของตัวแบบ โดยมีวัตถุประสงค์เพื่อสร้างตัวแบบที่สามารถพยากรณ์จำนวนวัสดุคงคลังให้ได้จำนวนที่ใกล้เคียงกับคำเฉลยเดิมหรือน้อยกว่าเดิมประสิทธิภาพในการพยากรณ์ของแต่ละตัวแบบ ดังภาพที่ 2 และภาพที่ 3



ภาพที่ 2 ตัวอย่างกราฟภาพรวมผลการพยากรณ์ค่าสูงสุด

เมื่อพิจารณาข้อมูลจากภาพที่ 2 และภาพที่ 3 พบว่า อัลกอริทึม Random forest เป็นอัลกอริทึมที่ให้ค่าการพยากรณ์ได้เท่ากับค่าเดิมมากที่สุดทั้งในส่วนของการพยากรณ์ค่าสูงสุดและค่าต่ำสุด อีกทั้งเมื่อพิจารณาความสามารถในการพยากรณ์ของตัวแบบจากอัลกอริทึม Random forest พบว่า สามารถพยากรณ์การจัดเก็บวัสดุได้เท่าเดิมและน้อยกว่าเดิม

ทั้งส่วนของค่าสูงสุดและค่าต่ำสุดถึงร้อยละ 78 (1,613 ระเบียน) และร้อยละ 83 (1,735 ระเบียน) จากทั้งหมด 2,077 ระเบียน ซึ่งแสดงให้เห็นว่าอัลกอริทึมนี้เป็นอัลกอริทึมที่เหมาะสมสำหรับนำมาใช้ในการพยากรณ์ เนื่องจากให้ค่าประสิทธิภาพที่เหมาะสมที่สุด เมื่อพิจารณาทั้งผลการพยากรณ์ข้อมูลใหม่และค่ารากที่สองของความคลาดเคลื่อน อีกทั้งสอดคล้องกับความต้องการของบริษัทที่อยากได้ตัวแบบที่สามารถลดจำนวนการจัดเก็บวัสดุในคลังให้มากที่สุด แต่เพียงพอต่อการใช้งานเมื่อต้องการ นอกจากนี้ยังพบว่า อัลกอริทึม Multi-layer perceptron พยากรณ์ค่าได้มากกว่าเดิมมากที่สุดประมาณ 2000 รายการ ในขณะที่อัลกอริทึม Random forest และอัลกอริทึม K-star สามารถพยากรณ์ค่าได้น้อยกว่าเดิมและเท่าเดิมในจำนวนที่ใกล้เคียงกัน



ภาพที่ 3 ตัวอย่างกราฟภาพรวมผลการพยากรณ์ค่าต่ำสุด

สำหรับการพยากรณ์ค่าต่ำสุด พบว่า อัลกอริทึม Random tree พยากรณ์ค่าได้มากกว่าเดิมจำนวนน้อยที่สุด ในขณะที่อัลกอริทึม Random forest และ K-star พยากรณ์ค่าได้น้อยกว่าเดิมและเท่าเดิมในจำนวนรายการที่ใกล้เคียงกัน ซึ่งในการที่อัลกอริทึมสามารถพยากรณ์ค่าได้น้อยกว่าเดิมหรือเท่าเดิมนั้นจะทำให้สามารถลดหรือไม่เพิ่มค่าใช้จ่ายในการจัดเก็บวัสดุอะไหล่ แต่ถ้าอัลกอริทึมที่พยากรณ์ค่าได้มากกว่าเดิม อาจจะทำให้ต้องเสียค่าใช้จ่ายในการจัดเก็บวัสดุอะไหล่ที่เพิ่มมากขึ้นไปด้วย ดังนั้นสำหรับกรณีศึกษานี้อัลกอริทึม Random forest จึงเหมาะสมที่จะมาใช้ในการพยากรณ์ปริมาณการจัดเก็บวัสดุอะไหล่

จากผลการทดลองทั้งหมด พบว่า ในด้านของประสิทธิภาพของอัลกอริทึม แม้ผลการทดลองไม่สอดคล้องกับงานวิจัยของ Rodrigo et al. (2017) ที่พบว่า อัลกอริทึม Random forest ไม่ใช่อัลกอริทึมที่ให้ผลการจำแนกที่ดีที่สุด ทั้งนี้อาจจะเนื่องมาจากความแตกต่างของข้อมูลทั้งด้านปริมาณและปัจจัย รวมถึงกระบวนการทดลอง อย่างไรก็ตามเมื่อพิจารณาถึงปัจจัยที่นำมาใช้ในการพยากรณ์ พบว่า ปัจจัยที่นำมาใช้ในการศึกษานี้กับปัจจัยในงานวิจัยของ Harshard & Priyanka (2015) นั้นมีความสอดคล้องกัน ซึ่งแสดงให้เห็นว่าปัจจัยดังกล่าวสามารถนำไปใช้ประยุกต์ใช้ในการพยากรณ์ค่าสูงสุดและค่าต่ำสุดในการจัดเก็บวัสดุอะไหล่ให้กับบริษัทอื่นๆ ที่มีลักษณะที่คล้ายคลึงกันได้ นอกจากนี้ในอนาคตเมื่อผู้วิจัยนำตัวแบบไปทดลองใช้โดยพัฒนาเป็นเว็บแอพพลิเคชัน เพื่อใช้ในฝ่ายงานจะสามารถลดค่าใช้จ่ายในการจัดเก็บวัสดุอะไหล่ของบริษัทได้ในอนาคต

สรุปผลการวิจัย

การสร้างตัวแบบในการพยากรณ์จำนวนการจัดเก็บวัสดุสินค้าในคลังแต่ละรายการ เพื่อหาจำนวนต่ำสุดและสูงสุดที่เหมาะสม โดยใช้เทคนิคการเรียนรู้ของเครื่องถูกนำมาใช้ในการวิเคราะห์ข้อมูลจริงจากบริษัทผลิตปูนซีเมนต์แห่งหนึ่ง โดยอัลกอริทึมที่ถูกนำมาใช้ในการศึกษาเปรียบเทียบเพื่อหาตัวแบบที่เหมาะสม ได้แก่ Random forest Multi-layer perceptron K-nearest neighbor K-star และ Random tree และวัดประสิทธิภาพโดยใช้ค่า RMSE โดยผลที่ได้ในขั้นตอนสร้างตัวแบบเพื่อพยากรณ์ค่าสูงสุด พบว่า อัลกอริทึม Random forest ให้ค่า RMSE ที่ดีที่สุด เท่ากับ 2.040 ส่วนผลที่ได้ในขั้นตอนการสร้างตัวแบบสำหรับการพยากรณ์ค่าต่ำสุด พบว่า อัลกอริทึม Random tree จะมี RMSE น้อยที่สุด เท่ากับ 0.803 นอกจากนี้ผลการวัดประสิทธิภาพของตัวแบบกับชุดข้อมูลใหม่เพื่อทดสอบประสิทธิภาพการพยากรณ์ พบว่า อัลกอริทึม Random forest เป็นอัลกอริทึมที่ให้ค่าประสิทธิภาพใกล้เคียงกับข้อมูลเดิมมากที่สุด ดังนั้นจึงอาจกล่าวได้ว่าสำหรับกรณีศึกษานี้อัลกอริทึม Random forest เป็นตัวแบบในการพยากรณ์จำนวนวัสดุอะไหล่ในคลังเพื่อช่วยให้ฝ่ายสามารถจัดการวัสดุคงคลังได้อย่างมีประสิทธิภาพมากขึ้น

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับความร่วมมือจากบริษัทผลิตปูนซีเมนต์แห่งหนึ่ง และได้รับเงินสนับสนุนในการนำเสนองานวิจัยจากคณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

เอกสารอ้างอิง

ฤทธิไกร ทวีเจริญ. (2559). **ตัวแบบการพยากรณ์การลาออกของพนักงานฝ่ายผลิตในธุรกิจเครื่องฉีดพลาสติกของบริษัทแห่งหนึ่ง**. (วิทยานิพนธ์ปริญญามหาบัณฑิต). มหาวิทยาลัยธรรมศาสตร์, คณะวิทยาลัยนวัตกรรม, สาขาวิชานโยบายและการบริหารเทคโนโลยีสารสนเทศ.

ทิตา ค้าผล. (2559). **การพยากรณ์กระแสไฟฟ้าขัดข้องโดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษาการไฟฟ้าส่วนภูมิภาค เขต 3 ภาคกลาง**. (วิทยานิพนธ์ปริญญามหาบัณฑิต). มหาวิทยาลัยธรรมศาสตร์, คณะวิทยาลัยนวัตกรรม, สาขาวิชานโยบายและการบริหารเทคโนโลยีสารสนเทศ.

บทเรียนออนไลน์ อัลกอริทึมการค้นหาเพื่อนบ้านใกล้สุด k ตัว. สืบค้นเมื่อ 22 ธันวาคม 2562, จากวิกิพีเดีย https://th.wikipedia.org/wiki/อัลกอริทึมการค้นหาเพื่อนบ้านใกล้สุด_k_ตัว.

บทเรียนออนไลน์ โครงข่ายประสาทเทียม. สืบค้นเมื่อ 17 เมษายน 2563, จากวิกิพีเดีย https://th.wikipedia.org/wiki/โครงข่ายประสาทเทียม.

พัฒนพงษ์ ดลรัตน์ และจารี ทองคำ. (2560). การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความเสร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ. **วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, 37**(3), 380–388.

วรางคณา เรียนสุทธิ์. (2016). การพยากรณ์จำนวนผู้ว่างงานในประเทศไทย. **Naresuan University Journal: Science and Technology, 24**(1), 102–114.

วิชญ์พงศ์ ดรุณธรรม. (2018). **บทเรียนออนไลน์ เจาะลึก Random Forest**. สืบค้นเมื่อ 15 พฤศจิกายน 2562, จาก https://medium.com/@witchapongdaroontham/.

เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล. (2558). การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา. **Science and Technology, 2**(2), 1–17.

อรรณพ กางกั้น. (2562). การพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซลโดยเทคนิควิธีเหมืองข้อมูล. **Science Journal Chandrakasem Rajabhat University, 29**(2), 39–47.

Harshad, S. & Priyanka V. (2015). Optimization of inventory level at Oilfield services: Case Study. **International Journal of Scientific & Engineering Research, 6**(4), 308–313.

Rodrigo, B. S., Eduardo, P. A. & Leonardo G. (2017**).** Predicting Material Backorders in Inventory Predicting Material Backorders in Inventory Management using Machine Learning, **Proceedings of the 4th IEEE Latin American Conference on Computational Intelligence** (pp. 1–6). DOI: 10.1109/LA–CCI.2017.8285684, Peru.

Sawarni, H. (2018). Improved Inventory Management Performance in Indonesia Spare–Parts Company Using ABC Classification and Min–Max Method. **Saudi Journal of Business and Management Studies (SJBMS), 3**(3), 248–252.

Yoan, M. J., Julio, M. Q. & Ireimis, L. V. (2016). Study of the Performance of the K* Algorithm in International Database. **Revista Politécnica, 12**(23), 51–56.