การทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้การเรียนรู้ของเครื่อง

กัญญณัช คุณาบุตร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตวิทยาลัย สาขาวิชาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2565

LABOR PRODUCTIVITY FORECASTING FOR JEWELRY SETTING USING MACHINE LEARNING

Kanyanat Khunabut

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology
Graduate School
Thai-Nichi Institute of Technology
Academic Year 2022

หัวข้อวิทยานิพนธ์	การทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้การเรียนรู้ของเครื่อง
โดย	กัญญณัช คุณาบุตร
สาขาวิชา	เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา	ดร.ณปภัช วิชัยดิษฐ์

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.พิชิต สุขเจริญพงษ์)
วันที่ ..... เดือน ..... พ.ศ. .....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ปริศนา มัชฌิมา)

.....กรรมการ
(ดร.ศรายุทธ นนท์ศิริ)

.....กรรมการ
(ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์
(ดร.ณปภัช วิชัยดิษฐ์)

กัญญณัช คุณาบุตร : การทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้การเรียนรู้ของเครื่อง. อาจารย์ที่ปรึกษา : ดร.ณปภัช วิชัยดิษฐ์, 62 หน้า.

ปัจจุบันหลายบริษัทในอุตสาหกรรมผลิตเครื่องประดับอัญมณีมีการบันทึกข้อมูลที่เกี่ยวข้องกับกระบวนการผลิตลงในระบบบริหารจัดการทรัพยากรภายในองค์กร (Enterprise resource planning system: ERP) เพื่อให้แต่ละแผนกสามารถเข้าถึงข้อมูลที่เกี่ยวข้องได้ รวมไปถึงการใช้งานระบบควบคู่ไปกับกระบวนการผลิตเพื่อเป็นบันทึกการรับ-จ่ายชิ้นงาน แต่ผู้วิจัยจบว่าหลายบริษัทยังไม่มีการนำข้อมูลที่ถูกบันทึกไว้ไปใช้ให้เกิดประโยชน์เท่าที่ควร ผู้วิจัยจึงมีแนวคิดในการนำข้อมูลการรับ-จ่ายตัวเรือน และอัญมณีของหน่วยงานฝังในปี พ.ศ. 2560-2564 มาใช้ในการศึกษาเปรียบเทียบประสิทธิภาพการทำนายค่าผลิตภาพแรงงานของแบบจำลองของเทคนิคที่สนใจ

ผู้วิจัยได้พัฒนาแบบจำลองการทำนายค่าผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อให้บริษัทสามารถทำนายค่าผลิตภาพ (Productivity) ในกระบวนการผลิตเครื่องประดับได้อย่างมีความแม่นยำมากยิ่งขึ้น ทำให้สามารถนำผลการทำนายไปใช้ประโยชน์ในการวางแผนการผลิต และเป็นตัวชี้วัดในการทำกิจกรรมปรับปรุงกระบวนการต่อไปได้ โดยผู้วิจัยใช้ข้อมูลที่ถูกบันทึกไว้ 5 ปีย้อนหลังในการทำการทดลองสร้างแบบจำลอง ใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree), เทคนิคป่าแบบสุ่ม (Random Forest), เทคนิคเพื่อนบ้านที่ใกล้ที่สุด k (K-Nearest Neighbors) และเทคนิคเกรเดียนท์บูสติ้ง (Gradient Boosting) ในการสร้างแบบจำลองการทำนาย

ผลวิจัยพบว่าการทำนายผลิตภาพแรงงานกระบวนการฝังอัญมณีโดยใช้เทคนิคป่าแบบสุ่ม (Random Forest) มีประสิทธิภาพมากที่สุด โดยเมื่อใช้ชุดข้อมูลที่มีระยะเวลา 4 ปี จะได้ค่า Adjusted R-squared เท่ากับ 0.5880 และค่า Root Mean Square Error (RMSE) เท่ากับ 0.1874 เมื่อนำไปทดลองใช้กับชุดข้อมูลทดสอบพบว่า สามารถทำนายค่าได้ใกล้เคียงความเป็นจริง มีค่า Adjusted R-squared เท่ากับ 0.5290 และค่า Root Mean Square Error (RMSE) เท่ากับ 0.2113 นอกจากนี้ยังพบว่าปัจจัยที่ส่งผลต่อค่าผลิตภาพแรงงานในกระบวนการฝังมากที่สุด 3 อันดับ คือ ทักษะของพนักงานฝังอัญมณี, ช่วงเวลา และประเภทการฝัง ตามลำดับ

บัณฑิตวิทยาลัย ลายมือชื่อนักศึกษา .....

สาขาวิชา เทคโนโลยีสารสนเทศ ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2565

KANYANAT KHUNABUT: LABOR PRODUCTIVITY FORECASTING FOR JEWELRY SETTING USING MACHINE LEARNING. ADVISOR: DR. NAPAPHAT VICHAIDIS, 62 PP.

Presently many companies in the jewelry manufacturing industry have been recording their manufacturing data in an enterprise resource planning system (ERP) for easier access to interesting data. Including using the system along with the production line as a work recording tool. The researcher found out that many companies didn't obtain the advantage from the recorded data as much as it supposed to. Therefore, the researcher has a concept to use the recorded data to study, develop the forecasting model from machine learning techniques and compare the performance of each model.

The researcher developed models to forecast labor productivity in jewelry settings using machine learning techniques so that the company was able to forecast labor productivity more precisely. the result of forecasting can be utilized in production planning and can be an indicator for further process improvement activities. The researchers used the recorded data from the past five years for experiments. Decision trees, Random Forest, K-Nearest Neighbors, and Gradient Boosting were used to develop the forecasting model.

The research revealed that the most suitable to forecast labor productivity in the jewelry setting process was Random Forest. With 4 years of training data, the model achieved Adjusted R-squared 0.5880 and Root Mean Square Error (RMSE) 0.1874. After applying these models to forecast labor productivity, Adjusted R-squared 0.5290 and RMSE 0.2133 were obtained. In addition, it was found that the three factors affecting labor productivity in the jewelry setting process the most were the skill of stone setter, seasonal, and setting type respectively.

Graduate School Student's Signature.....

Field of Study Information Technology Advisor's Signature.....

Academic Year 2022

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความกรุณา และการช่วยเหลือสนับสนุนอย่างยิ่งจาก ดร.ณปภัช วิชัยดิษฐ์ ที่ให้คำปรึกษาแนะนำที่เป็นประโยชน์ในการทำวิจัย และให้คำแนะนำในการปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเมตตา ด้วยความเอาใจใส่อย่างยิ่งตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้ และขอขอบคุณผู้บริหาร และพนักงานของบริษัท พาเทอร์สัน จิวเวลเลอรี่ จำกัด ที่ให้ความอนุเคราะห์ข้อมูลเพื่อนำมาใช้ในการวิจัย รวมไปถึงให้ความรู้ คำแนะนำที่เป็นประโยชน์เป็นอย่างยิ่ง

อีกทั้งผู้วิจัยตระหนักถึงความทุ่มเท และตั้งใจจริงของคณะกรรมการทุกท่าน ได้แก่ ผู้ช่วยศาสตราจารย์ ดร.ปริศนา มัชฌิมา ดร.ศรายุทธ นนท์ศิริ และ ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์ ที่กรุณาให้คำแนะนำที่เป็นประโยชน์ในการปรับปรุง แก้ไขในการทำวิทยานิพันธ์ฉบับนี้ เป็นอย่างดีจนสามารถทำให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์

นอกจากนี้ทางผู้วิจัยขอขอบคุณครอบครัว เพื่อนร่วมชั้นเรียน และบุคคลรอบตัวที่สนับสนุนให้คำแนะนำ และกำลังใจจนสามารถจัดทำวิทยานิพันธ์ฉบับนี้จนเสร็จสิ้น

ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะก่อให้เกิดประโยชน์แก่ผู้ที่สนใจ สามารถนำไปต่อยอดเพื่อประยุกต์ใช้กับอุตสาหกรรมที่เกี่ยวข้อง และนำไปใช้เป็นแนวทางในการเพิ่มประสิทธิภาพในกระบวนการผลิตต่อไป

กัญญณัช คุณาบุตร

สารบัญ

หน้า

สารบัญ (ต่อ)

สารบัญ (ต่อ)

สารบัญตาราง

ตาราง หน้า

สารบัญรูป

บทที่ 1
บทนำ

1.1 สภาวะความเป็นมา แนวทางเหตุผล และปัญหา

ในปี 2564 ที่ผ่านมา อุตสาหกรรมอัญมณีและเครื่องประดับเป็นอุตสาหกรรมที่มีมูลค่าการส่งออกมากเป็นอันดับที่ 5 ของประเทศไทย โดยมีมูลค่ารวม 317,888.61 ล้านบาท คิดเป็น 3.71% ของมูลค่าการส่งออกทั้งหมด จากการนำเสนอข้อมูลของสำนักงานปลัดกระทรวงพาณิชย์ [1] พบว่ามูลค่าการส่งออกเป็นรองอุตสาหกรมผลิตชิ้นส่วนยานยนต์ ชิ้นส่วนคอมพิวเตอร์ ผลิตภัณฑ์ยาง และเมล็ดพลาสติกตามลำดับ จึงเห็นได้ว่าอุตสาหกรรมอัญมณีและเครื่องประดับค่อนข้างมีความสำคัญกับเศรษฐกิจไทย ซึ่งปัจจุบันการผลิตเครื่องประดับในหลายขั้นตอนยังใช้แรงงานของมนุษย์ที่มีทักษะสูงเป็นส่วนใหญ่ แม้ว่าบางขั้นตอนสามารถนำเครื่องจักรบางอย่างมาใช้แทนที่ได้ เช่น การขัด หรือบางขั้นตอนที่สามารถนำเครื่องจักรมาช่วยทุ่นแรงในการผลิต โดยใช้บุคคลที่มีความเชี่ยวชาญในขั้นตอนนั้นๆ เป็นผู้ควบคุม แต่ก็ยังมีบางขั้นตอนก็ต้องอาศัยประสบการณ์ ทักษะ และความชำนาญของมนุษย์เพื่อรับมือกับข้อจำกัดทางธุรกิจบางอย่างในการผลิต เช่น ขั้นตอนการฝังอัญมณีลงบนตัวเรือน เป็นต้น

ขั้นตอนในการผลิตเครื่องประดับในอุตสาหกรรมนั้น ส่วนมากจะมีกระบวนการผลิตหลักใกล้เคียงขึ้น ซึ่งอาจมีความแตกต่างกันบ้างตามประเภทของเครื่องประดับ และการออกแบบ ดังนั้นทางผู้วิจัยจึงขอยกตัวอย่างขั้นตอนการผลิตเครื่องประดับแบบที่ไม่ซับซ้อนให้เห็นภาพมากยิ่งขึ้น โดยการทำชิ้นงานเทียนที่มีลักษณะเหมือนชิ้นงานจริง เพื่อประกอบเป็นต้นเทียน แล้วนำต้นเทียนดังกล่าวไปทำเป็นเบ้าปูนสำหรับเทน้ำโลหะเข้าไปในเบ้าปูน เมื่อทำการหล่อเรียบร้อยแล้วจะได้ออกมาเป็นชิ้นงานโลหะ แต่ผิวชิ้นงานหล่อได้ออกมามักไม่เรียบ, การแต่ง และประกอบตัวเรือน คือการนำชิ้นงานที่ได้จากการหล่อมาแต่งให้ได้ขนาด รูปทรงที่ต้องการ แล้วทำการประกอบให้เป็นชิ้นเดียวกันในกรณีที่ชิ้นงานนั้นๆมีส่วนประกอบมากกว่า 1 ชิ้น โดยจะใช้ความร้อนจากหัวเชื่อมแก๊ส หรือเครื่องเชื่อมเลเซอร์เป็นตัวให้ความร้อนกับชิ้นงาน และใช้ลวดเชื่อมเป็นตัวประสานเนื้อโลหะเข้าด้วยกัน, การฝังอัญมณีบนตัวเรือน คือการยึดอัญมณีเข้ากับตัวเรือนโลหะโดยใช้เทคนิคที่เหมาะสมกับตัวเรือนและอัญมณีชนิดนั้น เช่น การฝังหุ้ม การฝังหนามเตย การฝังจิกไข่ปลา และขั้นตอนสุดท้ายคือการขัดเงา เป็นขั้นตอนที่ทำการขัดผิวโลหะให้มีความเงางาม ในบางชิ้นงาน อาจนำไปทำการชุบเคลือบผิวเพื่อให้ได้สีโลหะตามที่ต้องการอีกด้วย

เมื่อกล่าวถึงอุตสาหกรรมการผลิตแล้ว มักจะพบการนำระบบบริหารจัดการทรัพยากรภายในองค์กร (Enterprise Resource Planning : ERP) มาใช้ในการบันทึกข้อมูลที่เกี่ยวข้องกับการผลิตควบคู่กันไปกับกระบวนการผลิต เช่น รายละเอียดของชิ้นงานที่ได้รับจากลูกค้า บันทึกการส่งคำสั่ง

การผลิต บันทึกการซื้อ การนำวัตถุดิบเข้ามาใช้ในการผลิต หรือบันทึกรายละเอียดของกระบวนการผลิตในแต่ละขั้นตอนซึ่งถือเป็นข้อมูลที่มีความสำคัญเพื่อนำไปสู่การนำข้อมูลทั้งหมดที่มีไปใช้ในการคำนวณต้นทุน ราคาขาย โดยบันทึกข้อมูลในฝ่ายผลิตมักจะเป็น วันและเวลาที่จ่ายให้-รับคืนจากพนักงานฝ่ายผลิต จำนวนชิ้น น้ำหนัก ส่วนประกอบเพิ่มเติมที่มีการจ่ายให้พนักงาน เช่น ลวดเชื่อม น้ำประสาน อัญมณี เปอร์เซ็นการอนุมัติให้สูญหาย รวมไปถึงชื่อของพนักงานที่เกี่ยวข้องกับชิ้นงานในแต่ละขั้นตอน โดยข้อมูลเหล่านี้สามารถนำไปเป็นข้อมูลตั้งต้นในการวางแผนการผลิต สร้างดัชนีชี้วัดการปรับปรุงกระบวนการ และสามารถนำไปใช้วิเคราะห์เพื่อตอบปัญหาทางธุรกิจได้

การวางแผนการผลิต คือการวางแผนการใช้ทรัพยากรที่มีอยู่ในองค์กร เช่น แรงงาน เครื่องจักร วัตถุดิบ กระบวนการผลิต (4M : Man, Machine, Material, Method) มาใช้ในการผลิตเพื่อให้ได้ผลลัพธ์ตามที่ต้องการ โดยเหตุผลที่บริษัทควรมีการวางแผนการผลิตก็เพื่อหาจุดสมดุลในการผลิต เพื่อที่จะได้ทรัพยากรให้น้อยที่สุด และให้ได้ผลผลิตมากที่สุด ก็จะทำให้บริษัทได้ผลกำไรกลับมามากขึ้น ซึ่งแผนการผลิตนั้นมีทั้งในระยะยาว และระยะสั้น โดยแผนการผลิตระยะยาวนั้นมักจะมีระยะเวลา 3-5 ปี มุ่งเน้นไปที่การขยายกิจการ การเพิ่มกำลังการผลิต เช่น ขยายพื้นที่โรงงาน การซื้อเครื่องจักรเพิ่ม การจ้างงานเพิ่มเพื่อให้สอดคล้องกับความต้องการของลูกค้าที่อาจจะเกิดขึ้นในอนาคต ส่วนแผนการผลิตระยะสั้น จะเป็นการวางแผนการผลิตในช่วงเวลาหนึ่งๆ เช่น แผนการผลิตประจำวัน แผนการผลิตประจำเดือน เป็นต้น โดยมักจะทำการเฝ้าติดตาม และควบคุมสถานการณ์ไปพร้อมๆกับการวางแผนการผลิต โดยเราสามารถวัดประสิทธิภาพของกระบวนการผลิตได้โดยการใช้ผลิตภาพ (Productivity) ซึ่งก็คืออัตราส่วนระหว่างผลผลิต (Output) และวัตถุดิบที่ป้อนเข้าไป (Input) ยกตัวอย่างเช่น จำนวนอัญมณีที่พนักงานคนหนึ่งสามารถฝังเข้ากับตัวเรือน (เม็ด) ต่อชั่วโมง เป็นต้น

ซึ่งผู้วิจัยพบว่าหลายบริษัทในอุตสาหกรรมอัญมณีและเครื่องประดับแม้จะมีการเก็บบันทึกข้อมูลไว้ก็จริง แต่ไม่ได้มีการนำข้อมูลมาใช้วิเคราะห์ให้เกิดประโยชน์มากเท่าที่ควร ดังนั้นผู้วิจัยจึงมีความสนใจที่จะนำข้อมูลที่ถูกบันทึกไว้ในระบบบริหารจัดการทรัพยากรภายในองค์กรของบริษัทมาใช้เป็นกรณีศึกษาเพื่อหาแนวทางที่เหมาะสมในการทำนายผลิตภาพ โดยใช้การเรียนรู้ของเครื่อง

1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์ข้อมูลที่ถูกสรุปเป็นแผนภาพข้อมูลกำลังการผลิตของแผนกฝังอัญมณี

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่สร้างโดยใช้เทคนิคการเรียนรู้ของเครื่องสำหรับการทำนายค่าผลิตภาพ (Productivity) ในกระบวนการฝังอัญมณีจากข้อมูลที่มีอยู่

1.2.3 เพื่อทราบถึงปัจจัยที่มีผลต่อค่าผลิตภาพการผลิต

1.3 ขอบเขตของงานวิจัย

1.3.1 ข้อมูลที่นำมาใช้ในงานวิจัยนี้ เป็นข้อมูลของแผนกฝังอัญมณี ของบริษัท พาเทอร์สัน จิวเวลเลอรี่ จำกัด ซึ่งข้อมูลดังกล่าวถูกบันทึก และรวบรวมไว้ในระบบบริหารจัดการทรัพยากรภายในองค์กร (Enterprise Resource Planning: ERP) ตั้งแต่ปี พ.ศ. 2560 จนถึงปี พ.ศ. 2564

1.3.2 ข้อมูลที่รวบรวมมาถูกบันทึกอยู่ในรูปแบบ .csv ซึ่งผู้วิจัยใช้โปรแกรม Microsoft Power BI ในการสร้างชุดแผนภาพเพื่อทำความเข้าใจข้อมูล และใช้โปรแกรมภาษาไพทอน (Python programming language) เป็นเครื่องมือในการสร้างแบบจำลอง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้ข้อมูลที่มีถูกต้อง เหมาะสมกับการนำไปใช้ในการทำนายผลิตภาพ

1.4.2 ได้กระบวนการในการทำนายผลิตภาพ และแบบจำลองที่มีความเหมาะสมกับลักษณะข้อมูลที่มีอยู่

1.4.3 ได้แนวคิดในการทำนายผลิตภาพที่เหมาะสมกับลักษณะข้อมูลที่มีอยู่ และสามารถนำไปประยุกต์ใช้กับข้อมูลของส่วนงานอื่นๆต่อไปได้ในอนาคต

1.5 แผนการดำเนินการวิจัย

ผู้วิจัยได้ทำการวางแผนการดำเนินการวิจัยไว้ดังตารางที่ 1.1

ตารางที่ 1.1 แผนการดำเนินการวิจัย

ขั้นตอนการดำเนินการ	ปี 2564		ปี 2565							
	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.
วิเคราะห์ปัญหางานวิจัย										
ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้อง										
รวบรวมข้อมูลที่ใช้ในการศึกษา และทำความสะอาดข้อมูล										
วิเคราะห์ข้อมูลเพื่อหาปัจจัยที่เกี่ยวข้อง										
ทดลองหาแบบจำลองที่เหมาะสมกับการทำนายผลิตภาพแรงงาน										
สรุปผลการวิจัย										

1.6 นิยามศัพท์เฉพาะ

1.6.1 การฝังอัญมณี (Stone Setting) หมายถึง การยึดอัญมณีให้ติดอยู่บนตัวเรือนโดยใช้การเครื่องมือโน้มส่วนหนึ่งของตัวเรือนที่เป็นโลหะมาเกาะกับขอบอัญมณี ซึ่งวิธีการฝังอัญมณีที่ได้รับความนิยมมีหลายรูปแบบ เช่น ฝังหุ้ม, ฝังหนามเตย, ฝังจิกไข่ปลา, ฝังไร้หนาม เป็นต้น ซึ่งความยากง่ายของการฝังแต่ละประเภทขึ้นอยู่กับหลายปัจจัยมาประกอบกัน เช่น ชนิดโลหะของตัวเรือน ประเภทของอัญมณี ขนาดของอัญมณี เป็นต้น

1.6.2 ระบบบริหารจัดการทรัพยากรภายในองค์กร (Enterprise Resource Planning: ERP) หมายถึง ระบบที่ประกอบด้วยระบบงานต่างๆของของค์กรไว้ภายในระบบเดียวกัน เช่น การจัดซื้อ การวางแผนการผลิต การจัดเก็บสินค้า การส่งออกสินค้า เป็นต้น เพื่อให้เกิดการเชื่อมโยงของข้อมูล และนำข้อมูลไปใช้ให้เกิดประโยชน์ได้อย่างสูงสุด

1.6.3 การทำนายผล (Forecasting) คือการนำข้อมูลที่มีอยู่มาใช้เป็นข้อมูลตั้งต้นในการทำนายค่าบางอย่างเพื่อนำไปใช้วางแผน หรือเป็นตัวเลขอ้างอิง ซึ่งสามารถทำนายได้อย่างถูกต้องแม่นยำเมื่อผู้ศึกษามีความเข้าใจในความเป็นมา ปัจจัยส่งผลให้ข้อมูลเกิดความเปลี่ยนแปลง ซึ่งวิธีในการทำนายผลข้อมูลนั้นปัจจุบันสามารถแบ่งออกเป็น 3 วิธี คือ การใช้สถิติ (Statistical) การใช้โปรแกรมจำลองสถานการณ์ (Simulation) ซึ่งสองวิธีการนี้มีข้อจำกัดบางอย่างที่ทำให้บางครั้งอาจจะไม่ได้ผลการทำนายที่แม่นยำเท่าไรนัก และอีกวิธีการหนึ่ง คือ การใช้การเรียนรู้ของเครื่อง (Machine Learning) เป็นวิธีการทำนายที่มีความน่าสนใจ สามารถรับมือกับความหลากหลายของข้อมูลได้ดีกว่า

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ขั้นตอนการฝังอัญมณี สามารถอธิบายได้ดังนี้

(1) ตรวจสอบตัวเรือนเบื้องต้น เช่น ตรวจสอบว่าบริเวณกระเปาะ (พื้นที่สำหรับวางอัญมณีบนตัวเรือน) สำหรับฝังถูกขัดเรียบร้อยแล้ว, รูปทรง หรือขนาดของหนามเตยที่หล่อมาพร้อมกับตัวเรือนมีความเหมาะสมกับอัญมณีที่จะนำมาฝัง สามารถฝังอัญมณีได้โดยไม่มีปัญหา ชิ้นงานก็จะถูกจ่ายไปให้พนักงานฝัง ในกรณีที่พบว่าตัวเรือนไม่ได้คุณภาพจะถูกส่งกลับไปยังกระบวนการก่อนหน้าเพื่อทำการแก้ไข หรือในกรณีที่เป็นตัวเรือนที่ลูกค้าส่งมาโดยเฉพาะ อาจจะแก้ปัญหาโดยการนำอัญมณีไปเจียรไนใหม่เพื่อลดขนาดลงเพื่อให้เข้ากับตัวเรือนได้

(2) เตรียมตัวเรือนสำหรับฝังอัญมณี โดยการยึดตัวเรือนให้อยู่กับที่ไว้กับครั่ง และใช้อุปกรณ์จับยึดส่วนที่เป็นครั่งไว้แทนที่จะยึดกับส่วนที่เป็นชิ้นงานเลย เพื่อป้องกันการเสียหายของชิ้นงานเมื่อได้รับแรงกด

(3) ฝังอัญมณี โดยวางอัญมณีลงบนจุดที่จะฝัง เช่น บนกระเปาะ หรือบริเวณหลุมที่เจาะไว้บนตัวเรือน จากนั้นใช้อุปกรณ์ในการดันหนามเตย หรือส่วนที่ใช้ยึดอัญมณีให้ติดอยู่กับตัวเรือน และทำการแต่งบริเวณที่ทำการฝังให้เรียบร้อย พนักงานจะทำการตรวจสอบคุณภาพเบื้องต้นก่อน เช่น อัญมณีแตกร้าวจากการรับแรงกดมากเกินไปหรือไม่, หน้าอัญมณีที่ฝังมีความเอียงหรือไม่ เป็นต้

(4) ล้างทำความสะอาดชิ้นงาน โดยพนักงานจะจุ่มล้างชิ้นงานลงในสารละลายเพื่อกำจัดครั่งออกจากชิ้นงาน เป่าชิ้นงานให้แห้ง

(5) ตรวจสอบคุณภาพ ชิ้นงานจะถูกตรวจสอบคุณภาพการฝังอย่างละเอียด เพื่อป้องกันปัญหาคุณภาพในกระบวนการฝัง เช่น อัญมณีขยับ, หน้าอัญมณีไม่เสมอกัน (ในกรณีที่ฝังเป็นแถวยาว), หนามเตยขนาดไม่เท่ากัน ไม่สมมาตร, ลักษณะโดยรวมไม่ตรงกับคำสั่งผลิต เป็นต้น [2]

2.1.2 ประเภทการฝังอัญมณี (Stone Setting Type)

การฝังอัญมณีสามารถแบ่งได้หลายประเภท ตามลักษณะของก้านโลหะที่ยึดอัญมณีให้ติดกับตัวเรือนได้ดังนี้

(1) การฝังหนามเตย (Prong setting) เป็นการฝังอัญมณีโดยการใช้ขาหรือก้านโลหะ (หรือที่เรียกว่าหนามเตย) เป็นตัวยึดเกาะขอบอัญมณีไว้ โดยก้านหนามเตยนั้นสามารถมีได้หลายลักษณะ ขึ้นอยู่กับการออกแบบ เช่น เป็นทรงกระบอก ทรงสี่เหลี่ยม หรือมีลักษณะคล้ายตัววี ที่เรียกว่าฝังหัวเรือ (V-Prong setting) ที่เหมาะสำหรับยึดเกาะอัญมณีที่มีมุมแหลม เช่น ทรงหยดน้ำ

โดยการฝังลักษณะนี้เหมาะกับการฝังอัญมณีเม็ดใหญ่ [3] ข้อดีของการฝังหนามเตยคือ เป็นวิธีการฝังที่ง่าย ไม่ซับซ้อน เสียเนื้อโลหะในการฝังน้อย ได้ชิ้นงานที่มีความโปร่ง น้ำหนักเบา สามารถออกแบบตรงส่วนของกระเปาะขึ้นไปจนถึงหนามเตยได้ค่อนข้างอิสระ แต่เนื่องจากเนื้อโลหะไม่ได้คลุมเนื้ออัญมณีทั้งหมด ดังนั้นอาจจะต้องสวมใส่ด้วยความระมัดระวัง

(2) การฝังไข่ปลา (Pavé setting) เป็นการฝังโดยการเจาะช่องบนตัวเรือน แล้ววางอัญมณีลงในช่องพร้อมกับจิกเนื้อโลหะขึ้นมายึดอัญมณีให้ติดกับตัวเรือน แล้วแต่งส่วนที่จิกขึ้นมาให้มีลักษณะกลมเงาเหมือนไข่ปลา เหมาะกับอัญมณีขนาดเล็ก และมีการเรียงกันจำนวนมากๆ [3] ซึ่งการฝังลักษณะนี้ใช้เวลาค่อนข้างมากเพราะพนักงานต้องจิกเนื้อโลหะขึ้นมาด้วยตัวเอง ภายหลังมาจึงมีการปรับปรุงให้การฝังลักษณะนี้ง่ายขึ้นโดยการทำไข่ปลาไว้บนแม่พิมพ์ เจาะช่องบนตัวเรือนไว้เบื้องต้น เพื่อให้หล่อออกมาแล้วตัวเรือนมีไข่ปลา และช่องสำหรับวางอัญมณีไว้แล้ว พนักงานฝังเพียงวางอัญมณีลงช่อง แล้วโน้มเนื้อโลหะมายึดอัญมณีได้เลย ทำให้การฝังลักษณะนี้ง่ายขึ้น และใช้เวลาน้อยลงมาก

(3) การฝังหุ้ม (Bezel setting) เป็นการฝังที่ใช้เนื้อโลหะหุ้มรอบๆขอบอัญมณี เป็นวิธีการฝังที่ใช้เนื้อโลหะมาก และใช้เวลานาน แต่การฝังประเภทนี้จะแข็งแรงกว่าการฝังประเภทอื่นๆ เพราะเนื้อโลหะคลุมรอบอัญมณีหมด ทำให้ไม่ได้รับความเสียหายหากเกิดการกระแทก [3]

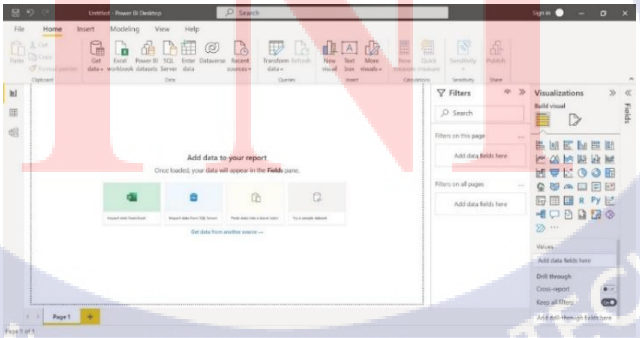
2.1.3 หลักการ data analysis and visualization

การแสดงผลของข้อมูลภาพ (Data visualization) คือการนำข้อมูลดิบที่ได้จากการเก็บรวบรวมบันทึกมาแสดงผลสรุปด้วยแผนภาพต่างๆ เช่น แผนภูมิแท่ง, แผนภูมิเส้น, แผนภูมิวงกลม หรือแผนภาพการกระจาย ซึ่งการจะเลือกใช้แผนภูมิแต่ละชนิดก็ขึ้นอยู่กับลักษณะของข้อมูลที่มี และสิ่งที่ต้องการสื่อให้ผู้ที่พบเห็นแผนภูมินั้นเข้าใจ ซึ่งการมองเห็นข้อมูลด้วยรูปภาพนั้นทำให้เข้าใจข้อมูลได้ง่าย และใช้เวลาน้อยกว่าการมองข้อมูลดิบที่เป็นตารางข้อมูล ดังคำกล่าวที่ว่า “รูปหนึ่งรูป มีค่าเทียบเท่าคำพูดนับพัน” ซึ่งหมายถึง การจะเข้าใจแผนภาพหนึ่งได้ ต้องใช้คำมาอธิบายจำนวนมาก เช่น วิธีการได้มาซึ่งข้อมูล, สาเหตุของการเลือกแผนภาพนั้นๆมาแสดงข้อมูล รวมไปถึงแผนภาพนี้สามารถสื่อถึงอะไรได้บ้าง เป็นต้น [4]

นอกจากนี้การแสดงผลข้อมูลด้วยภาพยังมีประโยชน์ในการวิเคราะห์ข้อมูลอีกด้วย เช่น ทำให้เห็นถึงแนวโน้มของข้อมูล, การรวมกลุ่มกันของข้อมูล, ทำให้พบข้อมูลที่ผิดปกติ เช่น อาจจะมีการบันทึกข้อมูลผิดพลาด หรืออาจเกิดจากปัจจัยบางอย่างที่เข้ามารบกวนขณะที่มีการเก็บข้อมูล, ทำให้เห็นถึงความสัมพันธ์กันของข้อมูล เช่น เหตุการณ์บางอย่าง มักจะส่งผลให้ข้อมูลเกิดการเปลี่ยนแปลงไปทางใดทางหนึ่งเสมอ เป็นต้น [5]

รูปที่ 0.1 Magic Quadrant for Analytics and Business Intelligence Platforms 2021

ปัจจุบันมีเครื่องมือที่ใช้ในการวิเคราะห์ และแสดงสามารถสร้างแผนภาพจากข้อมูลให้เลือกมากมาย ดังรูปที่ 2.1 จากการจัดลำดับโปรแกรมด้าน Analytics and Business Intelligence พบว่ามีโปรแกรมที่น่าสนใจมากมาย เช่น Microsoft Power BI, Tableau, Qlik, SAP, SSRS และอื่นๆ แต่ในงานวิจัยนี้ผู้วิจัยเลือกที่จะนำ Microsoft Power BI มาใช้ในการวิเคราะห์และแสดงผลข้อมูล เนื่องจากโปรแกรมดังกล่าวเป็นโปรแกรมที่ถูกจัดอันดับอยู่ในกลุ่มผู้นำ จึงถือว่าเป็นโปรแกรมที่มีความน่าสนใจ จากรูปที่ 2.2 เป็นรูปที่แสดงถึงหน้าโปรแกรม Microsoft Power BI จะเห็นว่าโปรแกรมดังกล่าวสามารถนำเข้าข้อมูลจากหลากหลายแหล่งมาวิเคราะห์ร่วมกันได้ เช่น ไฟล์ Microsoft Excel, CSV, SQL Server เป็นต้น และเมื่อนำเข้าข้อมูลมาแล้วก็สามารถเข้าสู่กระบวนการ ETL (Extract, Transform, Load) ข้อมูลเพื่อให้มีลักษณะพร้อมใช้งานได้ ไปจนถึงสามารถสร้างแผนภาพนำเสนอข้อมูลได้อย่างสวยงาม เมื่อผู้วิจัยได้ลองใช้งานโปรแกรมดังกล่าวก็พบว่า ใช้งานง่าย และไม่จำเป็นต้องใช้ความรู้ทางเทคนิคมากก็สามารถใช้งานให้เกิดผลลัพธ์ที่น่าพึงพอใจได้ และใช้เวลาในการศึกษา ทำความเข้าใจไม่มากจนเกินไป



รูปที่ 0.2 โปรแกรม Microsoft Power BI

2.1.4 หลักการของข้อมูลขนาดใหญ่ (big data)

ข้อมูลขนาดใหญ่ (Big Data) นั้นเคยถูกให้ความหมายไว้เมื่อช่วงปี ค.ศ. 1990 ว่าเป็นข้อมูลที่มีประมาณมากจนซอฟท์แวร์รุ่นเก่าไม่สามารถประมวลผลออกมาได้ จึงต้องใช้เทคโนโลยีสมัยใหม่มารองรับข้อมูลขนาดใหญ่นี้ ทำให้มีการนิยามความหมายของคำว่าข้อมูลขนาดใหญ่ (Big Data) ว่า เป็นเครื่องมือที่ใช้ในการจัดการกับข้อมูล เช่น ข้อมูลของบริษัท ข้อมูลลูกค้า ลักษณะของสินค้าที่ผลิต บันทึกการผลิตของสินค้าแต่ละชิ้น การส่งสัญญาณของเซนเซอร์ เป็นต้น ซึ่งข้อมูลที่มีอยู่ปริมาณมากนี้สามารถนำไปใช้ต่อยอดให้เกิดประโยชน์ในด้านต่างๆได้มากมาย เช่น นำไปใช้ในการวางแผนการทำงาน แผนการผลิต เพื่อช่วยในการตัดสินใจในการดำเนินธุรกิจ หรือช่วยเพิ่มโอกาส ช่วยหาช่องทางในการทำธุรกิจเพิ่มเติมได้ ซึ่งข้อมูลที่จะสามารถเรียกได้ว่าเป็น Big Data นั้นจำเป็นต้องมีลักษณะสำคัญตามหลัก 3Vs ดังนี้

ปริมาณของข้อมูล (Volume) จำเป็นต้องมีปริมาณมาก กล่าวได้ว่า เป็นข้อมูลที่ต้องใช้พื้นที่ในการจัดเก็บมากเป็นหน่วยเทระไบต์ (Terabyte) หรืออาจจะเป็นหน่วยเพตะไบต์ (Petabyte) เลยก็เป็นได้ เนื่องจากข้อมูลที่เก็บไว้นั้นเป็นข้อมูลที่อาจจะมีโครงสร้าง หรือไม่มีโครสร้างก็ได้ เช่น ตัวอักษร ตัวเลข รูปภาพ จำนวนการคลิก ลักษณะสัญญาณที่เก็บจากเซนเซอร์ เป็นต้น

ความเร็ว (Velocity) หมายถึง ความเร็วในการรับข้อมูล ความถี่ในการประมวลผลข้อมูลนั้นต้องมีความรวดเร็ว สอดคล้องกับปริมาณของข้อมูลที่เพิ่มขึ้นเกือบตลอดเวลา เช่น ข้อมูลการสนทนาที่เป็นรูปแบบตัวอักษร การบันทึกเสียง เป็นต้น

ความหลากหลาย (Variety) กล่าวคือ ความหลากหลายของประเภทของข้อมูลที่ทำการจัดเก็บ ช่วยให้สามารถนำข้อมูลไปวิเคราะห์ต่อยอดได้มากขึ้น เช่น ข้อมูลที่บันทึกเป็นตัวอักษร ข้อมูลรูปภาพ หรือข้อมูลเสียงสัญญาณที่บันทึกไว้

ซึ่งขั้นตอนการนำข้อมูลไปใช้งานนั้นก็เหมือนกับข้อมูลขนาดปกติทั่วไป ซึ่งก็คือ การจัดเก็บข้อมูล (Storage), การประมวลผลข้อมูล (Processing) เช่น การจัดกลุ่มข้อมูลที่มีความใกล้เคียงกัน เพื่อนำไปวิเคราะห์ (Analyst) ต่อ โดยวิเคราะห์หารูปแบบของความสัมพันธ์กันของข้อมูล การหาแนวโน้มของข้อมูล หรือนำเสนอข้อมูลที่เป็นประโยชน์ต่อการตัดสินใจในเชิงธุรกิจต่อไป โดยการนำเสนอนั้นมักจะนำเสนอออกมาในรูปแบบของแผนภูมิ

ยกตัวอย่างการนำเอาข้อมูลขนาดใหญ่มาใช้ประโยชน์ เช่น การนำข้อมูลขนาดใหญ่มาช่วยในการคาดการณ์ความต้องการของผู้บริโภค เพื่อนำไปใช้วางแผนการผลิตและพัฒนาสินค้าใหม่ การคาดการณ์ความผิดปกติของเครื่องจักร เพื่อให้สามารถซ่อมบำรุงได้ก่อนที่จะเกิดความชำรุดเสียหาย รวมไปถึงเป็นข้อมูลที่ช่วยสนับสนุนการตัดสินใจในการขับเคลื่อนองค์กรให้เป็นไปในทิศทางที่กำหนดไว้

2.1.5 หลักการของปัญญาประดิษฐ์ (Artificial Intelligence: AI)

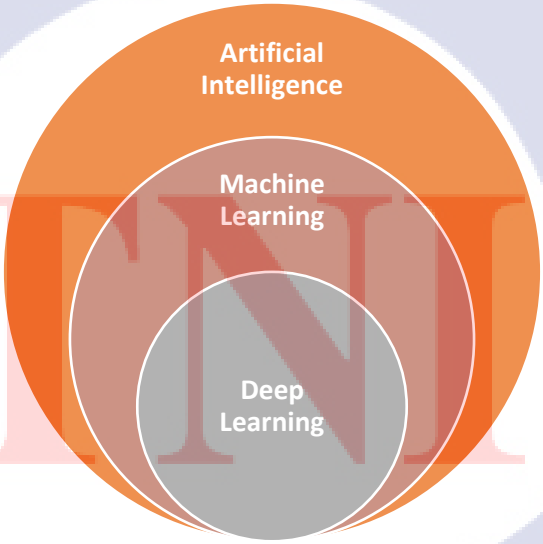
ปัญญาประดิษฐ์ หรือที่เรียกกันว่า AI นั้นเป็นเทคโนโลยีที่มนุษย์สร้างขึ้นมาเพื่อให้รับมือกับปัญหา และทำงานบางอย่างแทนมนุษย์ได้ โดยมีจุดมุ่งหมายในการนำเข้ามาใช้เพื่อลดต้นทุนและเพิ่มรายได้ให้กับองค์กร และสามารถแบ่งประเภทของปัญญาประดิษฐ์ได้ตามความสามารถ หรือความฉลาดของปัญญาประดิษฐ์ได้ดังนี้

ปัญญาประดิษฐ์เชิงแคบ (Narrow AI, Weak AI) คือปัญญาประดิษฐ์ที่มีความสามารถเฉพาะทางอย่างใดอย่างหนึ่งเหนือกว่ามนุษย์ เช่น AI ที่ช่วยในการผ่าตัด อาจจะมีความสามารถในการผ่าตัดมากกว่าหมอในปัจจุบัน แต่ไม่สามารถคิดเรื่องอื่นๆได้นอกจากการผ่าตัด โดย AI ชนิดนี้ถูกสร้างขึ้นมาเพื่อทำงานร่วมกับมนุษย์ ในแง่ของการช่วยตัดสินใจ ช่วยลดความซ้ำซ้อนในการทำงานของมนุษย์ ซึ่งปัจจุบัน AI ที่เกิดขึ้นส่วนมากมักจะเป็นประเภทนี้

ปัญญาประดิษฐ์ทั่วไป (General AI) คือปัญญาประดิษฐ์ที่มีความสามารถใกล้เคียงกับมนุษย์ สามารถทำกิจกรรมหลายๆอย่างได้เหมือนกับมนุษย์ทั่วไป

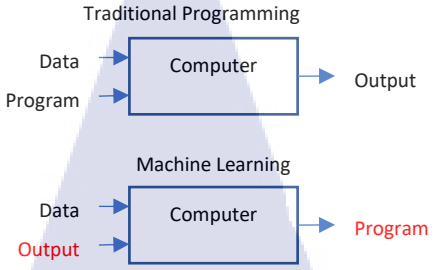
ปัญญาประดิษฐ์เชิงเข้ม (Strong AI) คือปัญญาประดิษฐ์ที่มีความสามารถสูงกว่ามนุษย์ปกติในหลายๆด้าน

นอกจากนี้ปัญญาประดิษฐ์ยังมีการแบ่งกลุ่มย่อยลงไปตามลักษณะการทำงานอีกด้วย ดังรูปที่ 2.3 จะมีการเรียนรู้ของเครื่อง (Machine Learning) และการเรียนรู้เชิงลึก (Deep Learning) เข้ามาเกี่ยวข้องด้วย



รูปที่ 0.3 การแบ่งกลุ่มย่อยของปัญญาประดิษฐ์

2.1.6 หลักการการเรียนรู้ของเครื่อง (Machine Learning: ML)



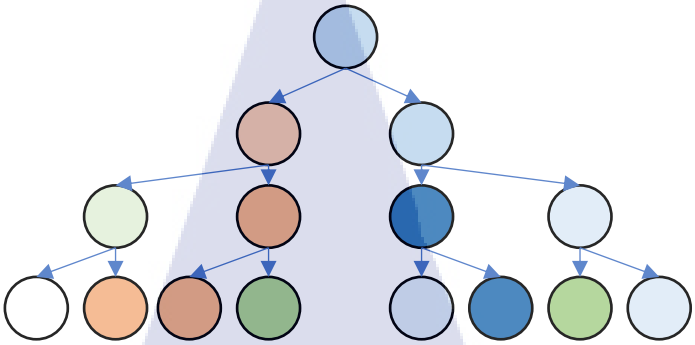
รูปที่ 0.4 เปรียบเทียบระหว่างการเขียนโปรแกรมแบบเดิม กับการใช้การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine Learning) เป็นกลุ่มย่อยของปัญญาประดิษฐ์ที่มีความแตกต่างจากการเขียนโปรแกรมทั่วไป เนื่องจากการเขียนโปรแกรมโดยทั่วไป จะใส่ข้อมูล (Data) กับวิธีคิดเข้าไป (Program) เพื่อให้ได้ออกมาเป็นผลลัพธ์ (Output) ดังรูปที่ 2.7 แต่การเรียนรู้ของเครื่อง จะเป็นการใส่ข้อมูลเข้าไปพร้อมกับผลลัพธ์ที่ต้องการ เพื่อให้คอมพิวเตอร์ทำการเรียนรู้ และสร้างวิธีคิดขึ้นมาเพื่อว่าในอนาคตมีข้อมูลใหม่ๆเข้ามาก็สามารถทำนายผลออกมาโดยใช้วิธีคิดที่มีอยู่แล้วได้เลย โดยการเรียนรู้ของเครื่องสามารถแบ่งได้ 3 แบบตามลักษณะการเรียนรู้ ซึ่งก็คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning), การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และการเรียนรู้จากสถานการณ์ (Reinforcement Learning)

การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การเรียนรู้ของเครื่องโดยใช้ข้อมูลตั้งต้นมาสอนให้รู้ว่า สิ่งนี้เรียกว่าอะไร สิ่งที่มีลักษณะข้อมูลแบบนี้ จัดอยู่ในกลุ่มใด เมื่อมีสิ่งนี้เกิดขึ้น อาจทำให้เกิดอีกสิ่งหนึ่ง เป็นต้น โดยสามารถแยกออกได้อีกเป็นสองประเภทย่อย คือ Classification เช่นการทำ Image Classification, Identity Fraud Detection และอีกกลุ่มคือ Regression เช่น การทำนายสภาพอากาศที่อาจจะเกิดขึ้น เมื่อมีปัจจัยต่างๆเข้ามาเกี่ยวข้อง, การพยากรณ์สภาวะการตลาด การทำนายการขยายตัวของกลุ่มประชากร ยกตัวอย่างแบบจำลองที่อยู่ในกลุ่ม Supervised Learning เช่น Support vector machine, Naïve Bayes, Decision Tree, Random Forest เป็นต้น

(1) เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

เป็นแบบจำลองทางคณิตศาสตร์ที่ใช้สำหรับการหาคำตอบที่ดีที่สุดตามรูปแบบ “ถ้าเงื่อนไข แล้วผลลัพธ์” สามารถแบ่งย่อยได้เป็น 2 ประเภท คือ Regression Tree สำหรับการวิเคราะห์การถดถอย (Regression Analysis) และ Classification Tree สำหรับการจำแนกหมวดหมู่ (Classification) สามารถเรียกรวมกันได้ว่า Classification And Regression Tree (CART) [6]



รูปที่ 0.5 ต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจแบบ Regression ประกอบด้วย โหนด (Node) หมายถึงคุณสมบัติต่างๆที่ใช้ในการตัดสินใจว่าจะไปในทิศทางใด, กิ่ง (Branch) หมายถึงคุณสมบัติในโหนดที่ถูกแยกออกมา และใบ (Leaf) คือกลุ่มของผลลัพธ์ที่เกิดจากการแบ่งแยกข้อมูลตามคุณสมบัติ โดยขั้นตอนของเทคนิคต้นไม้ตัดสินใจจะเริ่มจากการแบ่งข้อมูลออกเป็น 2 ส่วน (Recursive Binary Split) จากโหนดที่อยู่ล่างสุด แล้วหาจุดแบ่งข้อมูล (Split point) ที่เป็นไปได้ออกไปเรื่อยๆตามกิ่ง โดยจุดดังกล่าวมักจะเป็นจุดที่มีค่า Residual Sum of Squares (RSS) ต่ำที่สุด จนไปจบที่ Leaf node โดยการแบ่งข้อมูลแต่ละครั้งจะสามารถคำนวณค่า Residual Sum of Squares (RSS) ได้ดังสมการด้านล่าง

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{2.1}$$

RSS = residual sum of squares
y_i = i^th value of the variable to be predicted
$f(x_i)$ = predicted value of y_i
n = upper limit of summation

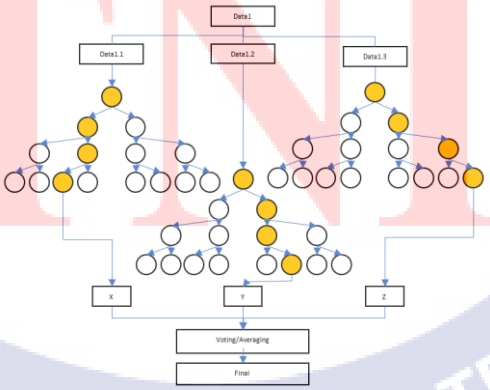
ส่วนเทคนิคต้นไม้ตัดสินใจแบบ Classification นั้นมีลักษณะเหมือนกับ Regression Tree แต่จะมีความแตกต่างกันตรงค่าที่ใช้ในการหาจุดแบ่ง ซึ่งจะใช้ Gini Impurity หรือ Entropy แทน เพื่อให้มีความเหมาะสมกับปัญหา โดย ค่า Gini Impurity คือ การวัดความไม่บริสุทธิ์ของกลุ่มข้อมูลที่ถูกแบ่งออกตามแต่ละจุดแบ่งข้อมูล (Split point) ในปัญหาการจัดกลุ่มแบบ Binary ที่มีเป้าหมายเป็น ซ้าย หรือขวา นั้นควรจะสามารถแบ่งกลุ่มออกได้อย่างชัดเจน เมื่อยิ่งแบ่งได้ออกมาชัดเจนเท่าไหร่ ก็ยิ่งทำให้ค่า Gini Impurity ลดน้อยลง ซึ่งสามารถคำนวณได้จากสมการด้านล่าง

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (2.2)$$

และค่า Entropy คือการวัดค่าความไม่แน่นอนของข้อมูล (Randomness) ที่เกิดขึ้นจากการที่ไม่สามารถคาดเดาผลของการทำนายได้ ดังนั้นในการทำ Classification จึงจำเป็นต้องลดความไม่แน่นอนที่อาจเกิดขึ้นให้ได้มากที่สุด ซึ่งค่า Entropy สามารถคำนวณได้จากสมการด้านล่าง

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk} \quad (2.3)$$

(2) เทคนิคป่าแบบสุ่ม (Random Forest) เป็นเทคนิคที่นำเอาเทคนิคต้นไม้ตัดสินใจ (Decision Tree) มาต่อยอด โดยการนำเอา Decision Tree จำนวนมากมารวมกันตั้งแต่ 10 แบบจำลอง หรืออาจมากกว่า 1,000 แบบจำลอง ซึ่งแต่ละ Decision Tree จะได้รับชุดข้อมูลย่อยที่มีความแตกต่างกันกันออกไป แล้วทำนายผลของแต่ละ Decision Tree ออกมา ถ้าเป็น Classification จะเลือกผลลัพธ์ที่ถูกเลือกโดย Decision Tree มากที่สุด และถ้าเป็น Regression จะให้ค่าเฉลี่ยของทุกๆผลลัพธ์ออกมาดังรูปที่ 2.6 ซึ่งเทคนิคป่าแบบสุ่มนี้เป็นเทคนิคที่ถูกสร้างขึ้นเพื่อลดข้อผิดพลาดของเทคนิคต้นไม้ตัดสินใจ ทำให้มีความแม่นยำกว่าเทคนิคต้นไม้ตัดสินใจ โดยเทคนิคป่าแบบสุ่มนั้นสามารถแบ่งได้เป็น 2 ประเภท คือ Bagging และ Boosting โดย Bagging คือการแยกกันเรียนรู้แล้วทำนายผลข้อมูลออกมา ส่วน Boosting คือ การเรียนรู้ แล้วปรับปรุงไปเรื่อยๆ [7]



รูปที่ 0.6 ประเภทของเทคนิคป่าแบบสุ่ม

(3) เทคนิค Gradient Boosting เป็นเทคนิคการเรียนรู้ของเครื่องที่มีลักษณะคล้ายกับเทคนิคป่าแบบสุ่ม (Random Forest) ที่เป็นการเรียนรู้จากการนำแบบจำลองที่มีการเรียนรู้แบบอ่อนมารวมเข้าด้วยกัน (Ensemble) ซึ่ง Gradient Boosting เป็นการเรียนรู้แบบนำเทคนิคต้นไม้ตัดสินใจมาเรียงต่อกัน เพื่อปรับปรุงประสิทธิภาพให้ดีขึ้น ซึ่งเทคนิค Gradient Boosting มี 2 เทคนิคย่อยที่มีความแตกต่างกัน ได้แก่ Extreme Gradient Boosting (XGBoost) เป็นเทคนิคที่นำ Decision Tree จำนวนมากมาต่อกัน เพื่อศึกษาข้อผิดพลาดของต้นก่อนหน้า เพื่อปรับปรุงให้ดีขึ้น และหยุดการทำงานเมื่อไม่พบข้อผิดพลาดของต้นก่อนหน้าให้แก้ไขแล้ว ส่วน Light Gradient Boosting (LightGBM) เป็นเทคนิคที่มีประสิทธิภาพสูง ใช้ การเรียนรู้แบบ Leaf-wise ที่สามารถลดการสูญเสียได้มากกว่าแบบ Level-wise ทำให้ได้ผลลัพธ์ที่แม่นยำ และมีประสิทธิภาพมากกว่า [8]

(4) เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เป็นเทคนิคที่ใช้ในการจำแนกข้อมูล (Classification) ที่สามารถใช้ได้ทั้งการจำแนกข้อมูล และการถดถอยของข้อมูล โดยอ้างอิงจากข้อมูลที่อยู่ใกล้ที่สุดจำนวน k ตัว โดยมีขั้นตอนดังนี้

กำหนดค่า k และระยะห่างระหว่างข้อมูลแต่ละตัว

คำนวณระยะห่างระหว่างชุดข้อมูลใหม่ และชุดข้อมูลเดิม

จัดลำดับความใกล้เคียง และเลือกขึ้นมาเป็นจำนวน k ตัว ที่อยู่ใกล้กัน

รวบรวมกลุ่มเป้าหมายของเพื่อนบ้าน

จัดกลุ่มให้ข้อมูลใหม่ โดยพิจารณาให้เป็นกลุ่มเดียวกับกลุ่มของเพื่อนบ้านที่มีจำนวนซ้ำมากที่สุดในข้อมูลที่เลือกมา k ตัว [9]

โดยสามารถคำนวณระยะห่างระหว่างข้อมูลได้หลายวิธีดังตารางที่ 2.1

ตารางที่ 0.1 แสดงวิธีคำนวณระยะห่างระหว่างข้อมูล

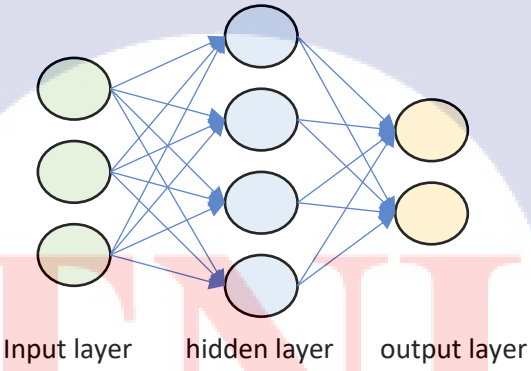
ชื่อวิธีคำนวณ	สมการ	ความหมายสมการ
1. ระยะห่างยูคลิเดียน (Euclidean Distance)	$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$ (2.4)	วัดระยะห่างระหว่างจุดโดยใช้ทฤษฎีพีทาโกรัส
2. ระยะห่างแมนฮัตตัน (Manhattan Distance)	$d(p, q) = \Sigma \lvert pi - qi \rvert$ (2.5)	วัดระยะห่างระว่างจุดโดยมีลักษณะการวัดแบบผังเมือง แมนฮัตตัน คือขยับขึ้นตามบล็อกสี่เหลี่ยมจตุรัส
3. Cosine Similarity	$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$ (2.6)	วัดระยะห่างระหว่างจุดด้วย Cosine

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือ การเรียนรู้ของเครื่องโดยที่ไม่มีการนำเข้าข้อมูลเพื่อสอน โดยแบบจำลองจะให้ความสำคัญที่การจัดกลุ่มข้อมูลที่มีลักษณะใกล้เคียงกันเอาไว้ด้วยกัน (Clustering) เช่น การจัดกลุ่มผู้บริโภคเพื่อแนะนำผลิตภัณฑ์ที่เหมาะสมกับลูกค้ากลุ่มนั้นๆ

การเรียนรู้จากสถานการณ์ (Reinforcement Learning) คือ การเรียนรู้จากสถานการณ์ที่เกิดขึ้นว่าถ้าเจอสถานการณ์นี้ แล้วทำแบบนี้ จะได้รางวัล แต่ถ้าทำไม่ถูกต้องจะถูกทำโทษซ้ำๆจะทำให้แบบจำลองสามารถจดจำได้ว่า ต้องทำอย่างไรในสถานการณ์ที่พบเจอ เหมาะกับการหากลยุทธ์เพื่อเอาชนะเกม

2.1.7 หลักการการเรียนรู้เชิงลึก (Deep Learning: DL)

การเรียนรู้เชิงลึกเป็นส่วนย่อยอีกหนึ่งส่วนของการเรียนรู้ของเครื่องที่มีการเลียนแบบการทำงานของเซลล์ประสาทในสมองมนุษย์ โดยมีการนำข้อมูลหลายๆข้อมูลเข้าไป แล้วประมวลผลออกมาเป็นคำตอบ ยกตัวอย่างเช่น เมื่อสายตามนุษย์มองเห็นแมว รูปของแมวที่มนุษย์มองเห็นจะถูกนำไปประมวลผล โดยที่ก็ไม่รู้ว่าประมวลผลอย่างไร แต่ก็ได้คำตอบกลับมาว่า นี่คือแมว โดยเปรียบเทียบตัวอย่างนี้ กับรูปที่ 2.7 สามารถมองได้ว่า รูปแมวที่มองเห็นคือ Input Layer วิธีการประมวลผลที่ไม่รู้ว่าประมวลผลอย่างไร คือ Hidden Layer และผลลัพธ์ที่ทราบว่าเป็นแมว คือ Output Layer



รูปที่ 0.7 หลักการของ Deep Learning

ดังนั้นการเรียนรู้เชิงลึกก็คือ การที่มี Hidden Layer หลายชั้น เพื่อให้สามารถคิดได้อย่างซับซ้อนมากขึ้น คล้ายกับสมองของมนุษย์นั่นเอง ดังนั้นเมื่อเปรียบเทียบระหว่างการเรียนรู้ของเครื่อง กับการเรียนรู้เชิงลึกแล้ว จะพบว่าเมื่อมีปริมาณข้อมูลมากๆการเรียนรู้เชิงลึกจะสามารถทำการประเมินได้เองว่าจะใช้ข้อมูลไหน และใช้แบบจำลองใดบ้างโดยที่ไม่ต้องพึ่งพามนุษย์นั่นเอง

2.1.8 การประเมินค่าประสิทธิภาพการทำงานของแบบจำลอง

ประสิทธิภาพของแบบจำลองนั้นสามารถประเมินได้จากหลากหลายค่า การเลือกใช้นั้นขึ้นอยู่กับลักษณะของข้อมูลที่นำมาสร้างแบบจำลองว่าเป็นข้อมูลลักษณะใด ถ้าข้อมูลเป็น Regression สามารถเลือกใช้ MAE, MSE, RMSE, R2, Adjusted-R2 หรือถ้าข้อมูลเป็น Classification สามารถเลือกใช้ Confusion Matrix หรือ ROC score ได้ เป็นต้น

(1) รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Square Error: RMSE) เป็นค่าที่นิยมใช้ในการประเมินประสิทธิภาพของแบบจำลองประเภท Regression [7] สามารถคำนวณได้ตามสมการด้านล่าง

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{2.7}$$

(2) ค่าสัมประสิทธิ์การกำหนด (Coefficient of Determination: R2) ใช้ในการอธิบายความสัมพันธ์ระหว่างตัวแปรต้น และตัวแปรตามที่ถูกเลือกมาใช้ในแบบจำลอง ถ้าตัวแปรมีความสัมพันธ์กันมากค่าสัมประสิทธิ์การกำหนดจะมีค่าเข้าใกล้ 1 ถ้ามีค่าน้อย หรือเข้าใกล้ 0 แปลว่าตัวแปรที่เลือกมาอาจมีความสัมพันธ์กันน้อย สามารถคำนวณได้ดังสมการด้านล่าง

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \tag{2.8}$$

(3) ค่า Adjusted-R Square เป็นค่าที่ใช้บ่งบอกว่าตัวแปรที่ที่ใส่เข้ามาในแบบจำลองนั้นมีประโยชน์ต่อการวิเคราะห์หรือไม่ ซึ่งจะมีค่าน้อยกว่าค่าสัมประสิทธิ์การกำหนดเสมอ ค่า Adjusted-R Square สามารถเพิ่มขึ้นได้ เมื่อใส่ตัวแปรที่มีผลต่อการวิเคราะห์เข้าไปในแบบจำลอง ในทางกลับกัน ถ้าใส่ตัวแปรที่ไม่มีประโยชน์เข้าไปก็ทำให้ค่า Adjusted-R Square ลดลงได้ ในขณะที่ไม่ว่าจะใส่ตัวแปรที่ดีหรือไม่ดีเข้าไปในแบบจำลอง ค่าสัมประสิทธิ์การกำหนดก็ยังเพิ่มขึ้น ดังนั้นจึงนิยมใช้ค่า Adjusted-R Square มากกว่า สามารถคำนวณได้ตามสมการด้านล่าง

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1} \tag{2.9}$$

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 <u>การทำนายผลข้อมูล (Forecasting algorithm)</u>

โดยทั่วไปแล้ว จะสามารถทำนายข้อมูลได้เมื่อมีข้อมูลที่เกี่ยวข้องปริมาณมากเพียงพอที่จะนำมาใช้วิเคราะห์ และยังต้องมีความเข้าใจในข้อมูลที่จะนำมาวิเคราะห์มาก เพียงพอจึงจะสามารถทำนายข้อมูลได้อย่างมีประสิทธิภาพ ปัจจุบันมีวิธีการทำนายข้อมูล 3 ประเภท สามารถแบ่งได้ตามเครื่องมือที่ใช้ดังนี้ [10]

(1) เครื่องมือทางสถิติ (Statistical method) มาใช้ในการทำนาย โดยการวิเคราะห์การถดถอย (Regression Analysis) เป็นเครื่องมือที่ได้รับความนิยมเป็นอย่างมาก เช่น การนำ Multiple linear regression มาใช้ในการทำนายค่าผลิตภาพแรงงานในอุตสาหกรรมก่อสร้าง โดยวิธีการนี้จะนำปัจจัยต่างๆที่คาดว่าจะมีผลเข้าไปพิจารณาด้วย แต่เครื่องมือทางสถิติก็ยังมีข้อจำกัดอยู่ตรงที่สามารถนำเข้าปัจจัยที่นำเข้าไปพิจารณาได้อย่างจำกัด และการที่มนุษย์ทำงานเดิมซ้ำๆเป็นระยะเวลานานก็มักจะเกิดการเรียนรู้ (Learning curve) เทคนิคใหม่ๆเพื่อทำให้งานนั้นมีคุณภาพดีขึ้นใช้เวลาน้อยลง ดังนั้นการใช้ Regression Analysis จึงอาจยังไม่ใช่วิธีที่ดีที่สุดสำหรับการทำนายข้อมูลในปัจจุบัน [11]

(2) โปรแกรมจำลองสถานการณ์ (Simulation program) เช่น การนำโปรแกรมจำลองสถานการณ์ที่มีชื่อว่า System dynamics เข้ามาทดลองใช้ในการทำนายข้อมูล โดยผู้ใช้งานต้องนำเข้าข้อมูลปัจจุบันเข้าไปใช้ในการศึกษาก่อน เช่น จำนวนเครื่องจักร, จำนวนพนักงาน, เวลาที่ใช้ในการทำงานของแต่ละกระบวนการ รวมไปถึงเป้าหมายที่ต้องการจะได้รับจากการจำลองสถานการณ์ ซึ่งโปรแกรมมีข้อดีส่วนที่สามารถเข้าใจถึงโอกาสที่จะเกิดเหตุการณ์บางอย่างที่ไม่คาดคิด ซึ่งเคยมีการนำทฤษฎีการจัดสมดุลการผลิต (Line Balancing method) เข้ามาใช้ในการจำลองสถานการณ์ในกระบวนการหล่อเครื่องประดับเพื่อลดการว่างของพนักงานที่เกิดขึ้นจากการรอคอยชิ้นงานจากกระบวนการก่อนหน้าในกระบวนการหล่อเครื่องประดับ ซึ่งมีสมมติฐานว่าเมื่อมีการปรับปรุงจุดดังกล่าวแล้วจะสามารถเพิ่มประสิทธิภาพของกระบวนการหล่อเครื่องประดับได้มากขึ้น และปริมาณการผลิตต่อวันของหน่วยงานจะเป็นไปตามเป้าหมายที่วางไว้ โดยการปรับปรุงกระบวนการนั้นหมายถึงการใส่เครื่องจักร และพนักงานเพิ่มเข้าไปในส่วนที่เป็นคอขวด (Bottleneck) ของกระบวนการผลิต และเมื่อเพิ่มเครื่องจักร และพนักงานเข้าไปตามที่โปรแกรมจำลองสถานการณ์ได้ประเมินไว้แล้วพบว่า ประสิทธิภาพในการผลิตเพิ่มขึ้น มีความใกล้เคียงกับค่าที่โปรแกรมจำลองสถานการณ์ได้ทำนายไว้ [12]

(3) เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เป็นวิธีที่มีการศึกษาเพิ่มขึ้นในภายหลังเพื่อนำมาปรับใช้ในการทำนายข้อมูล โดยการนำเทคนิคการเรียนรู้ของเครื่องมาสร้างแบบจำลองโดยใช้ชุดข้อมูลตัวอย่างในการเรียนรู้ แล้วสามารถทำนายข้อมูลออกมาได้ เป็นวิธีที่สามารถใช้ข้อมูลตั้งต้นได้หลากหลายประเภท เช่น ข้อมูลตัวเลข, ข้อมูลตัวอักษร, ข้อมูลรูปภาพ และ

ข้อมูลประเภทอื่นๆ และมีการนำเทคนิคการเรียนรู้ของเครื่องมาทำนายข้อมูลในอุตสาหกรรมอย่างหลากหลายดังเช่นตารางที่ 2.2

ตารางที่ 0.2 งานวิจัยอื่นๆที่นำเทคนิคการเรียนรู้ของเครื่องมาใช้ทำนายในอุตสาหกรรม

Research's name	Industry section	Model	Source
Modeling and Forecasting of Producer Price Index (PPI) of Cheese Manufacturing Industries	Cheese manufacturing	LSTM, BI-LSTM, GRU	[13]
Supervised vs. unsupervised learning for construction crew productivity prediction	Construction industry	FFBP, GRNN, SOM	[14]
Hybrid Artificial Intelligence HFS-RF-PSO Model for Construction Labor Productivity Prediction and Optimization	Construction industry	ANN, RF	[10]
Simulation-based construction productivity forecast using Neural-Network-Driven Fuzzy Reasoning	Construction industry	ANN, NNDFR	[11] [15]
Regional Manufacturing Industry Demand Forecasting: A Deep Learning Approach	Manufacturing industry	LSTM, SVM, BP, AR, RF	[16]

จากตัวอย่างการทำนายผลิตภาพแรงงานในกระบวนการเทคอนกรีตของอุตสาหกรรมก่อสร้าง พบว่าปัจจัยที่มีผลต่อค่าผลิตภาพในกระบวนการเทคอนกรีตสามารถแบ่งได้เป็น 3 ประเภทหลัก คือ สภาพอากาศในแต่ละวัน, ลักษณะของพนักงานในโครงการ และลักษณะของพื้นที่ก่อสร้าง เมื่อ Farid Mirahadi, Emad Elwakil, Tarek Zayed นำข้อมูลการทำงานในขั้นตอนการเทคอนกรีตที่มีการเก็บบันทึกไว้มาจัดกลุ่มปัจจัยที่ส่งผลต่อค่าผลิตภาพแรงงานของขั้นตอนการเทคอนกรีตโดยการใช้ Subtractive clustering เป็นตัวจัดกลุ่มเบื้องต้น และใช้ K-means เป็นตัวยืนยัน

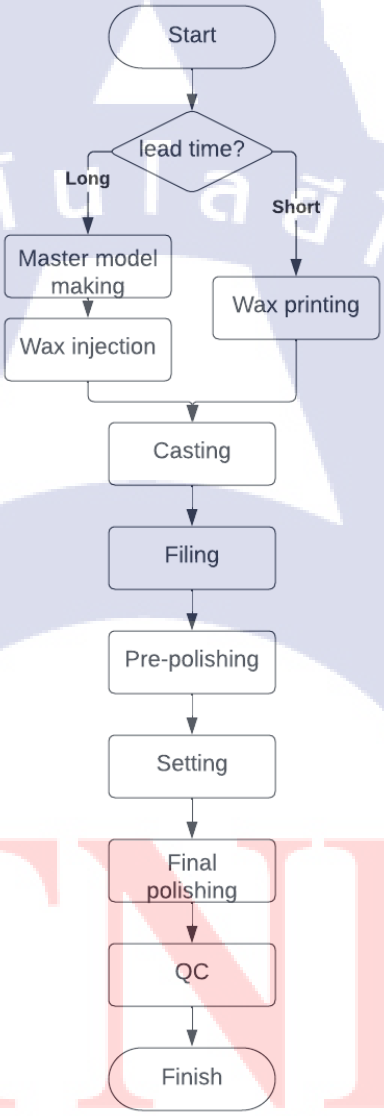
การจัดกลุ่มปัจจัยที่มีผลต่อค่าผลิตภาพ [15] แล้วหลังจากนั้นก็ทำการแบ่งข้อมูลที่เตรียมไว้เป็นสองส่วนเพื่อให้แบบจำลองเรียนรู้ และใช้ในการทดสอบประสิทธิภาพของแบบจำลอง ซึ่งแบบจำลองที่ถูกนำมาเปรียบเทียบในงานวิจัยนี้ คือ Artificial neural network (ANN), Adaptive-based fuzzy inference system (ANFIS), Neural network driven fuzzy reasoning (NNDFR) จากการทดสอบพบว่า จำนวนกลุ่มปัจจัยที่เหมาะสมที่สุดควรเป็น 3 กลุ่ม และใช้แบบจำลอง NNDFR เพื่อให้ได้ค่า Mean squared error (MSE) ต่ำที่สุด และเมื่อแบบจำลองมีค่า MSE ต่ำที่สุดก็แปลว่ามีความผิดพลาดลดลง สามารถทำนายค่าผลิตภาพแรงงานได้ใกล้เคียงกับความเป็นจริงมากที่สุด [11] [15]

นอกจาก ANN, ANFIS และ NNDFR ที่ถูกนำมาใช้ในการศึกษาเพื่อพัฒนาแบบจำลองที่ใช้ในการทำนายผลิตภาพแรงงานในอุตสาหกรรมก่อสร้างแล้ว Sara Ebrahimi, Aminah Robinson Fayek ก็ได้นำแบบจำลองอื่นๆมาทดสอบเพิ่มอีก ซึ่งข้อมูลที่นำมาใช้ในการศึกษาเป็นกระบวนการเทคอนกรีต ใช้เวลาในการเก็บข้อมูล 92 วัน พบว่ามีปัจจัยทั้งหมด 112 ปัจจัยที่มีผลกับค่าผลิตภาพแรงงาน ข้อมูลที่เก็บบันทึกมาถูกนำไปปรับปรุงเพื่อให้เหมาะกับการศึกษาโดยการ Normalization แปลงจากข้อมูลตัวอักษรเป็นตัวเลข เติมข้อมูลที่มีการเว้นว่างไว้ด้วยค่าเฉลี่ยของข้อมูล และกำจัดข้อมูลที่เป็นค่าผิดปกติออกจากชุดข้อมูล โดยงานวิจัยนี้นำแบบจำลอง Random Forest (RF) เข้ามาทำการทดลองเปรียบเทียบกับ ANN, ANFIS และ ANFIS-GA และทำการวัดผลประสิทธิภาพของแบบจำลองโดยใช้ค่า Accuracy ซึ่งได้ผลการทดลองว่า ANN ได้ค่า Accuracy ต่ำที่สุด ถัดมาเป็น ANFIS ที่ไม่มีการ Optimization โดยการใช้ Genetic algorithm (GA) ถัดมาเป็น ANFIS-GA และแบบจำลองที่มีประสิทธิภาพสูงสุดคือ Random forest (RF) ซึ่งสาเหตุที่ RF มีประสิทธิภาพสูงสุดเนื่องจากแบบจำลองนี้มีการสุ่มที่สามารถช่วยลดข้อผิดพลาดขนาดใหญ่ได้ [10]

2.2.2 <u>ปัจจัยที่มีผลต่อเวลาที่ใช้ในกระบวนการฝังอัญมณี</u>

(1) คุณภาพของชิ้นงานที่มาจากกระบวนการผลิตก่อนหน้า โดยผู้วิจัยจะกล่าวถึงกระบวนการผลิตเครื่องประดับอัญมณีอย่างสั้น และอธิบายว่าแต่ละกระบวนการส่งผลอย่างไรต่อกระบวนการฝังอัญมณี ตามรูปที่ 2.8 โดยกระบวนการผลิตเริ่มจากการทำชิ้นงานเทียนขึ้นมา ซึ่งปัจจุบันสามารถทำได้โดยการฉีดเทียนร้อนเข้าแม่พิมพ์ยาง (Wax injection) หรือการพิมพ์ด้วยเครื่องพิมพ์สามมิติ (3D wax printing) ซึ่งการฉีดเทียนนั้นเกิดขึ้นมาก่อน และมีการใช้กระบวนการนี้กันมาเป็นเวลากว่า 80 ปี [17] โดยลักษณะ ความสมบูรณ์ของชิ้นงานเทียนนั้นมีผลต่อกระบวนการถัดไปอย่างมาก ดังนั้นจึงจำเป็นต้องทำแม่พิมพ์ให้ดี ต้องมีการคำนวณค่าเผื่อการหดตัวของน้ำโลหะที่เหมาะสมเข้ามาร่วมด้วย เพราะถ้าไม่มีการบวกค่าเผื่อสำหรับการหดตัวของโลหะแล้ว จะทำให้ต้องใช้เวลาในการเทียบหาขนาดอัญมณีกับขนาดตัวเรือนที่เหมาะสมกันเพิ่มขึ้น [18]

และในช่วง 10 ปีที่ผ่านมาก็เริ่มมีการนำเทคโนโลยีใหม่เข้ามาใช้ในการลดเวลาการผลิต [17] โดยการออกแบบในโปรแกรม (Computer-aided design : CAD) แล้วพิมพ์ออกมาเป็นชิ้นงานเทียนโดยใช้เครื่องพิมพ์สามมิติซึ่งสามารถออกแบบชิ้นงานพร้อมกับคำนวณค่าเผื่อสำหรับการหดตัวของโลหะได้เลย ซึ่งการใช้ CAD คำนวณค่าเผื่อไว้ตั้งแต่ต้นกระบวนการผลิตสามารถลดเวลาที่ใช้ในขั้นตอนการฝังอัญมณีได้ประมาณ 70% ของเวลาที่ใช้อยู่เดิม [18] [19]



รูปที่ 0.8 ขั้นตอนการผลิตเครื่องประดับอัญมณี

เมื่อได้ชิ้นงานเทียนมาแล้ว จึงนำชิ้นงานไปประกอบขึ้นมาเป็นต้นเทียน เพื่อเข้าสู่กระบวนการหล่อ (Lost-wax casting) ซึ่งเป็นการเปลี่ยนสภาพต้นเทียนให้เป็นโลหะโดยการใช้ความ

ร้อนหลอมละลายโลหะแล้วเทลงในเบ้าปูนภายในสภาวะที่เหมาะสม เมื่อได้ต้นโลหะมาแล้วจึงทำการตัดชิ้นงานออกจากต้นโลหะ ทำการแต่งตัวเรือน (Filing) ให้ตัวเรือนมีรูปทรง ลักษณะที่ถูกต้องตามที่กำหนด ซึ่งถ้าแต่งตรงกระเปาะสำหรับวางอัญมณีได้เหมาะสมแล้ว ขั้นตอนการฝังจะสามารถใช้เวลาน้อยลง ขั้นตอนต่อมาคือการนำไปเข้ากระบวนการขัดหยาบ (Pre-polishing) เพื่อให้ได้ผิวชิ้นงานที่มีความเรียบมากขึ้น และต้องขัดบริเวณกระเปาะฝังให้เรียบ จึงจะเข้าสู่กระบวนการฝังอัญมณี (Stone setting) และการขัดเงา (Final polishing) เพื่อให้ชิ้นงานมีความเงางาม นอกจากนั้นอาจจะมีกระบวนการอื่น ๆ เพิ่มเติมตามแต่ละประเภทของเครื่องประดับ เช่น การประกอบสร้อย ชิ้นส่วนสำเร็จเข้ากับตัวเรือน การชุบเคลือบผิวเพื่อเพิ่มคุณสมบัติให้กับผิวชิ้นงาน

ดังนั้นจึงเห็นได้ว่า คุณภาพของชิ้นงานที่ส่งต่อกันไปในแต่ละกระบวนการมีความสำคัญมาก กล่าวได้ว่า ถ้าต้นทางผลิตมาดี ปลายทางก็ทำงานสะดวก

(2) ลักษณะทางกายภาพของชิ้นงาน สามารถแบ่งได้เป็นสองส่วน คือ ความแข็งของตัวเรือน และความแข็งของอัญมณี โดยโลหะที่มีความแข็งต่ำนั้นจะฝังอัญมณีได้ง่าย เพราะพนักงานไม่จำเป็นต้องใช้แรงในการดันเนื้อโลหะให้ไปในทิศทางที่ต้องการ เช่น การดันหนามเดยให้ไปเกาะหน้าอัญมณี โดยสามารถจัดลำดับความยาก-ง่ายได้เป็น แพลตินั่มที่มีความแข็งมากกว่า รองลงมาเป็นทอง ซึ่งทองแต่ละสูตรนั้นมีความแข็งต่างกันไปขึ้นอยู่กับปริมาณส่วนผสมของโลหะชนิดอื่น ๆ และเงินเป็นโลหะที่ฝังง่ายที่สุด [20]

คุณสมบัติอีกประการคือ ความแข็งของอัญมณี พบว่าอัญมณีที่ฝังง่ายที่สุด คือ อัญมณีที่มีค่าความแข็งสูงตามการจัดอันดับของ Moh’s hardness scale [21] เช่น เพชร ทับทิม ไพลิน และคอรันดัมสีอื่น ๆ เนื่องจาก อัญมณีเหล่านี้สามารถรับแรงกระแทกจากเครื่องมือขณะที่ฝังอัญมณีได้มากกว่า ทนต่อการขูดขีดได้มากกว่า รองลงมาเป็นอัญมณีสังเคราะห์ และอัญมณีที่มีความแข็งต่ำ เช่น มรกต โทปาซ ควอตซ์ ทัวร์มาลีน ฯลฯ นอกจากนี้ยังต้องคำนึงถึงแรงกดขณะที่ฝังอัญมณีในมุมที่ไม่เหมาะสมก็อาจจะทำให้อัญมณีเกิดความเสียหายได้ [22]

นอกจากปัจจัยหลักแล้ว สิ่งที่สำคัญไม่แพ้กันคือ ทักษะ ความชำนาญของพนักงานแต่ละคนที่เกิดจากการฝึกฝน เก็บประสบการณ์มาพัฒนาความสามารถของตนอย่างสม่ำเสมอ รวมไปถึงอุปกรณ์ที่ใช้ในการปฏิบัติงาน สภาพแวดล้อมในการทำงานของพนักงานที่ดีล้วนมีส่วนช่วยให้พนักงานปฏิบัติงานได้อย่างมีประสิทธิภาพยิ่งขึ้น

2.2.3 <u>ค่าผลิตภาพในอุตสาหกรรมการผลิต</u>

ค่าผลิตภาพ (Productivity) ถูกใช้เป็นดัชนีวัดผลการทำงานของบุคคล หรือหน่วยงาน โดยสามารถแบ่งได้เป็น 2 ประเภท คือ ผลิตภาพแรงงาน (Labor Productivity) และผลิตภาพรวม (Total Factor Productivity) โดยผลิตภาพแรงงานนั้นเป็นตัวเลขที่สามารถใช้วัดผลการทำงานของแรงงานมนุษย์ได้ สามารถคำนวณได้จาก

$$\text{Productivity} = \frac{\boldsymbol{Output}}{\boldsymbol{Input}} \tag{2.10}$$

โดยที่ผลลัพธ์ (Output) คือผลผลิตที่ได้จากการปฏิบัติงานเมื่อใช้ทรัพยากรจำนวนหนึ่ง เช่น จำนวนชิ้นงานที่แผนกหล่อสามารถหล่อได้ในแต่ละวัน จำนวนอัญมณีที่แผนกฝังสามารถฝังลงบนตัวเรือนได้ในแต่ละสัปดาห์ เป็นต้น ส่วนข้อมูลนำเข้า (Input) คือทรัพยากรที่ใส่เข้าไปในการผลิตเพื่อให้ได้ออกมาเป็นผลผลิต เช่น เวลา วัตถุดิบ จำนวนคน เป็นต้น ส่วนผลิตภาพรวม คือการนับรวมส่วนที่เกี่ยวข้องทั้งหมดในการผลิตเป็นข้อมูลนำเข้า เช่น ค่าแรง ค่าไฟ ค่าวัตถุดิบอุปกรณ์ ค่าสถานที่ ฯลฯ โดยค่าผลิตภาพนั้นเป็นเพียงค่าคงที่ในช่วงเวลาหนึ่ง สามารถเพิ่ม-ลดได้เมื่อมีปัจจัยภายนอกเข้ามาเกี่ยวข้อง ซึ่งสามารถจัดกลุ่มปัจจัยดังกล่าวได้ตามหลักการ 4M (Man, Machine, Material, Method) [23] ดังตารางที่ 2.3

ตารางที่ 0.3 ปัจจัยที่มีผลต่อค่าผลิตภาพในอุตสาหกรรมการผลิต

กลุ่ม	ปัจจัย
Man	สภาพแวดล้อมในการทำงาน ปัญหาส่วนตัว ปัญหาสุขภาพของพนักงาน การสนับสนุน การสอนงานที่เหมาะสมจากหัวหน้างาน
Machine	เครื่องจักร อุปกรณ์ที่มีมาตรฐานเหมาะสม
Material	วัตถุดิบที่มีคุณภาพ และมีการสำรองไว้เพียงพอ
Method	เทคโนโลยีที่ช่วยให้บุคคลภายนอกรบกวนฝ่ายผลิตน้อยลง ช่วยทำงานได้สะดวกมากขึ้น

ค่าผลิตภาพนั้นสามารถนำไปใช้ประโยชน์เป็นข้อมูลตั้งต้น หรือตัวชี้วัดของกิจกรรมอื่นได้หลากหลาย เช่น เป็นค่าที่นำมาใช้ประมาณในการวางแผนการผลิต การส่งมอบสินค้า เช่น รับรายการสินค้าที่ลูกค้าต้องการเข้ามาจำนวนชิ้นเท่านี้ จำเป็นต้องใช้เวลา หรือทรัพยากรอื่นเท่าไหร่เพื่อให้ใบสั่งผลิตดังกล่าวเสร็จลุล่วงตามเวลาที่กำหนด หรือใช้เป็นตัวชี้วัดในการทำกิจกรรมเพิ่มประสิทธิภาพของฝ่ายผลิตได้ เช่น กิจกรรม LEAN เป็นต้น [24]

โดยค่าผลิตภาพนี้สามารถเพิ่มขึ้นได้เมื่อมีการนำกิจกรรมบางอย่างเข้ามาปรับใช้กับแต่ละหน่วยงานในองค์กรดังตารางที่ 2.4

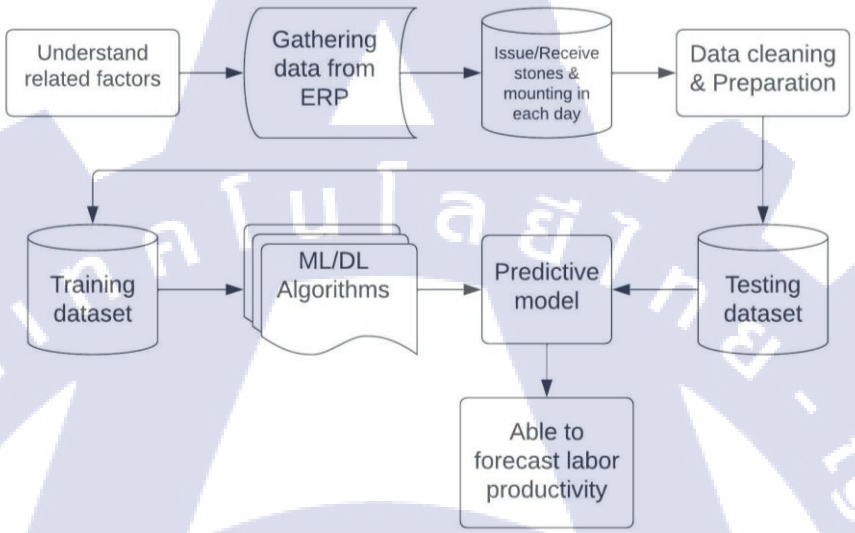
ตารางที่ 0.4 วิธีการเพิ่มผลิตภาพในองค์กร

วิธีการ	รายละเอียด
ลงทุนเพิ่ม	เพิ่มคน, เพิ่มเครื่องจักร
นำนวัตกรรมเข้ามาประยุกต์ใช้	ลดขั้นตอนในการทำงาน, ลดเวลาที่ใช้ในการทำงาน, เพิ่มคุณภาพในการผลิต และควบคุมคุณภาพ
คุณภาพขององค์กร	สร้างสภาพแวดล้อมในการทำงานที่ดี
จัดกิจกรรมแข่งขัน	เพื่อท้าทายความสามารถของบุคลากร
เพิ่มความสามารถ	ส่งเสริมให้มีการอบรมเพื่อเพิ่มทักษะของบุคลากร

โดยแต่ละวิธีการก็ใช้ทรัพยากรเข้ามาเป็นตัวช่วยในการเพิ่มผลิตภาพได้มากน้อยแตกต่างกันไป โดยส่วนใหญ่แล้วมักต้องการให้ทรัพยากรในการผลิตเท่าเดิม หรือน้อยลง แต่ได้ผลผลิตมากขึ้นจึงจะถือว่าประสบความสำเร็จในการทำกิจกรรมเพิ่มค่าผลิตภาพภายในองค์กร

บทที่ 3
วิธีดำเนินการงานวิจัย

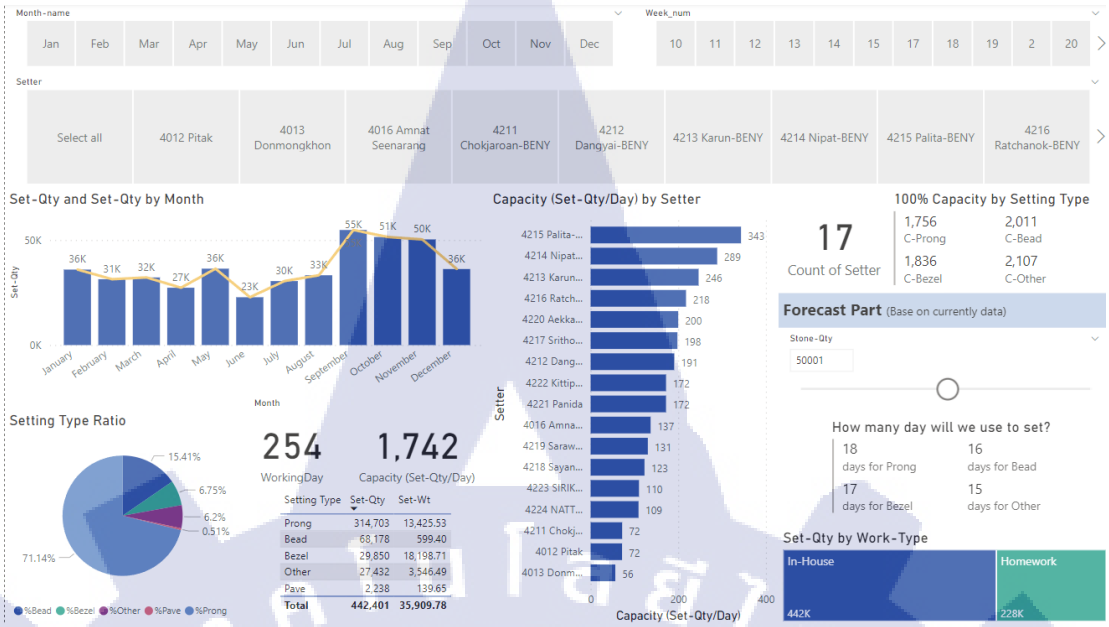
งานวิจัยเรื่อง การทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยการใช้การเรียนรู้ของเครื่อง ผู้วิจัยได้แบ่งขั้นตอนการดำเนินงานได้ดังรูปที่ 3.1



รูปที่ 0.1 ขั้นตอนการดำเนินการวิจัย

3.1 วิเคราะห์ปัญหางานวิจัย

ผู้วิจัยพบว่าบริษัท พาเทอร์สัน จิวเวลเลอรี่ จำกัด มีการใช้งาน และบันทึกข้อมูลการผลิตลงในระบบบริหารจัดการทรัพยากรภายในองค์กรมาเป็นระยะเวลาหลายปี แต่ไม่มีการนำข้อมูลที่บันทึกไว้มาใช้วิเคราะห์ให้เกิดประโยชน์ และยังไม่มีตัวชี้วัดประสิทธิภาพการทำงานของพนักงานฝ่ายผลิตที่ชัดเจน ผู้วิจัยจึงได้นำข้อมูลการผลิตของแผนกฝังอัญมณีในปี พ.ศ. 2564 มาสร้างเป็นแผนภูมินำเสนอดังรูปที่ 3.2 เมื่อทำการคำนวณผลิตภาพแรงงานจากค่าเฉลี่ยแล้วพบว่า ผลทำนายจากการเฉลี่ยนั้นไม่ได้ให้ความแม่นยำเท่าที่ควร จึงมีความสนใจในการนำวิธีการทำนายผลิตภาพด้วยเครื่องมืออื่นๆมาประยุกต์ใช้กับข้อมูลที่มีอยู่



รูปที่ 0.2 แผนภาพสรุปข้อมูลกำลังการผลิตของแผนกฝังอัญมณี ในปี พ.ศ. 2564

3.2 ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้อง

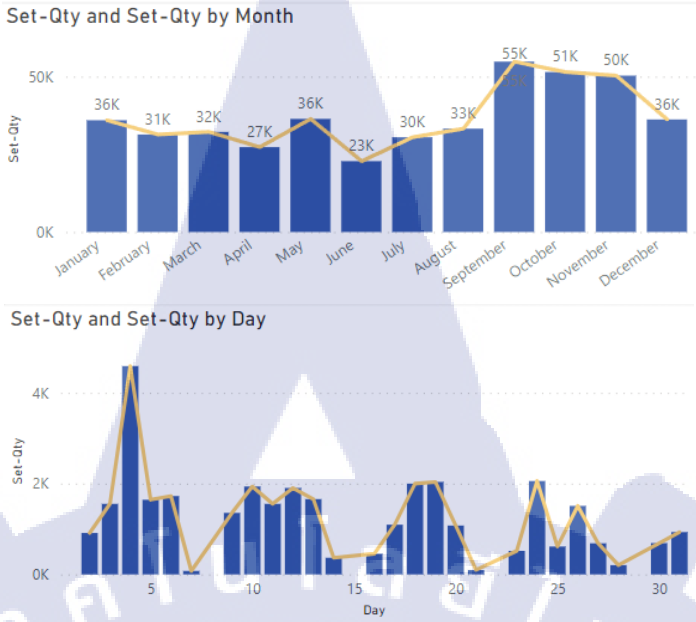
ผู้วิจัยทำการศึกษาทฤษฎีเกี่ยวกับการฝังอัญมณี เพื่อให้เข้าใจถึงความแตกต่างของการฝังอัญมณีแต่ละประเภท รวมถึงการไปศึกษาหน้างานเพื่อให้ทราบถึงปัจจัยที่คาดว่าจะมีผลในการฝังอัญมณีแต่ละประเภทด้วยเช่นกัน การทำนายค่าผลิตภาพด้วยวิธีต่างๆ การใช้แบบจำลองในการทำนายผลิตภาพแรงงานของอุตสาหกรรมที่มีความใกล้เคียงกัน มีลักษณะข้อมูลตั้งต้นที่ใกล้เคียงกันที่น่าจะมีความเป็นไปได้ที่จะนำมาปรับใช้กับข้อมูลที่ใช้ในการศึกษา

3.3 วิเคราะห์ข้อมูลเพื่อหาปัจจัยที่ที่มีผลกับค่าผลิตภาพแรงงาน

ผู้วิจัยทำการวิเคราะห์ข้อมูลที่ถูกสรุปเป็นแผนภาพข้อมูลกำลังการผลิตของแผนกฝังอัญมณี ในปี พ.ศ. 2564 ดังรูปที่ 3.2 เพื่อให้มีความเข้าใจในลักษณะของข้อมูลมากยิ่งขึ้น ซึ่งผู้วิจัยพบปัจจัยต่างๆดังนี้

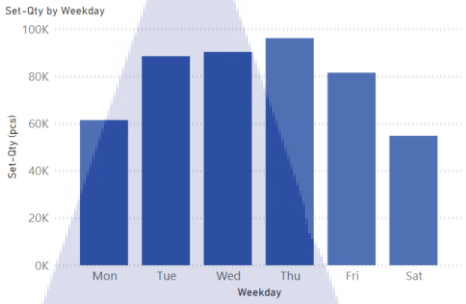
3.3.1 <u>ช่วงเทศกาลสำคัญประจำปี</u>

โดยในแต่ละเดือนของปี พ.ศ. 2564 ปริมาณการฝังอัญมณีนั้นมีแนวโน้มเพิ่ม-ลดตามเทศกาลสำคัญประจำปี เนื่องจากทางบริษัทไม่มีหน้าร้าน ไม่มีตราสินค้าเป็นของบริษัท แต่รับผลิตและส่งออกให้ลูกค้า นอกจากนี้ยังมีนโยบายผลิตตามคำสั่งผลิต (Made to order) ไม่มีการสำรองสินค้าในคลัง ทำให้ลูกค้ามักส่งคำสั่งผลิตเข้ามาก่อนช่วงเทศกาลสำคัญ เพื่อให้บริษัทสามารถผลิตและส่งสินค้าไปถึงลูกค้าก่อนถึงช่วงเทศกาล เพื่อที่ลูกค้าจะได้มีสินค้าสำหรับวางหน้าร้าน หรือขายได้



รูปที่ 0.3 แผนภาพแสดงปริมาณการฝังอัญมณีในแต่ละเดือนของปี พ.ศ. 2564 (บน) และในแต่ละวันของเดือนสิงหาคม (ล่าง)

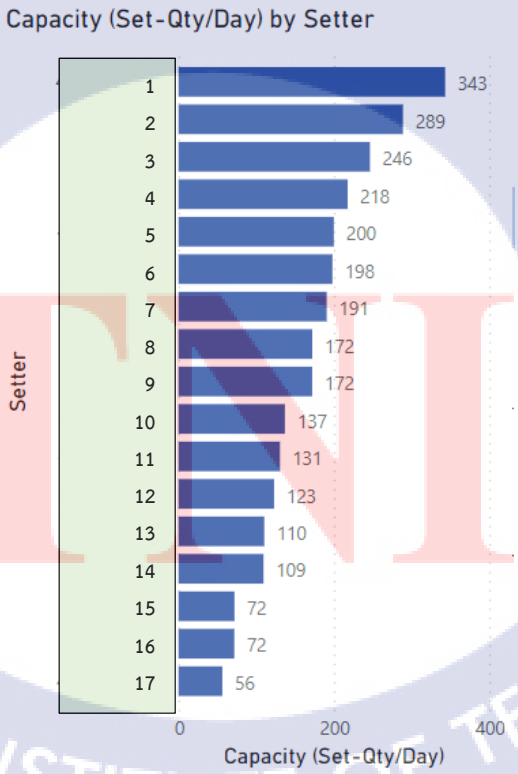
อย่างเพียงพอในช่วงเทศกาลสำคัญ เช่น ช่วงเดือนกันยายน-พฤศจิกายน ซึ่งเป็นช่วงก่อนเข้าสู่เทศกาลคริสต์มาส และปีใหม่จะเป็นช่วงที่มีปริมาณการผลิตสูงที่สุด ส่วนเดือนธันวาคม และเดือนเมษายนเป็นเดือนที่มีวันทำงานน้อย เนื่องจากเป็นช่วงปีใหม่ และสงกรานต์ ทำให้มีการรับคำสั่งผลิตจากลูกค้าลดลง และปริมาณการฝังอัญมณีต่อเดือนลดลงดังรูปที่ 3.3 ด้านบนและเมื่อแสดงผลข้อมูลปริมาณการฝังอัญมณีในแต่ละวันตามรูปที่ 3.3 ด้านล่างพบว่ามักมีการขึ้นลงของกราฟซ้ำๆ ในหน่วยสัปดาห์ แม้ว่าจะมีบางช่วงที่ไม่เป็นไปตามที่กล่าวไว้ เนื่องจากมีงานด่วนแทรกเข้ามาบ้างในบางช่วง และเมื่อทำการสรุปปริมาณการฝังอัญมณีตามวันทำงานดังรูปที่ 3.4 พบว่าแต่ละวันในสัปดาห์จะมีปริมาณการฝังอัญมณีแตกต่างกันไป เนื่องจากทางบริษัทมีการกำหนดวันส่งออกเป็นวันศุกร์ของทุกสัปดาห์ ทำให้กระบวนการฝังอัญมณีที่เป็นกระบวนการเกือบท้ายสุดจึงมักจะได้รับงานในช่วงก่อนวันส่งออกเสมอ ดังนั้นการวัดประสิทธิภาพการทำงาน การวางแผนในการจ่ายงาน รวมไปถึงการทำนายค่าจึงควรใช้หน่วยสัปดาห์เพื่อให้สอดคล้องกับลักษณะการทำงานของบริษัท



รูปที่ 0.4 แผนภาพแสดงปริมาณการฝังอัญมณีในแต่ละวันของปี พ.ศ. 2564

3.3.2 <u>ประสบการณ์ ทักษะ ความสามารถของพนักงาน</u>

เนื่องจากการฝังอัญมณีนั้นเป็นงานฝีมือที่อาศัยความชำนาญที่เกิดจากการฝึกฝนของพนักงาน เมื่อพนักงานได้รับงานที่มีลักษณะเดิมซ้ำๆก็มักจะเกิดการเรียนรู้ ทำให้เกิดเป็นเทคนิคส่วนตัวเพื่อให้สามารถทำงานได้สะดวกขึ้น ใช้เวลาน้อยลง และได้งานที่มีคุณภาพมากขึ้น ซึ่งพนักงานแต่ละคนก็จะมีเทคนิคที่ไม่เหมือนกัน จากรูปที่ 3.5 แสดงให้เห็นถึงประสิทธิภาพการฝังอัญมณีรวมทุกประเภทของพนักงานแต่ละคน พบว่าพนักงานทุกคนมีความแตกต่างกันอย่างชัดเจน ดังนั้นในการนำข้อมูลไปใช้ในการทำนาย ควรพิจารณาแยกเป็นรายบุคคลมากกว่าการมองเป็นภาพรวมทั้งแผนก

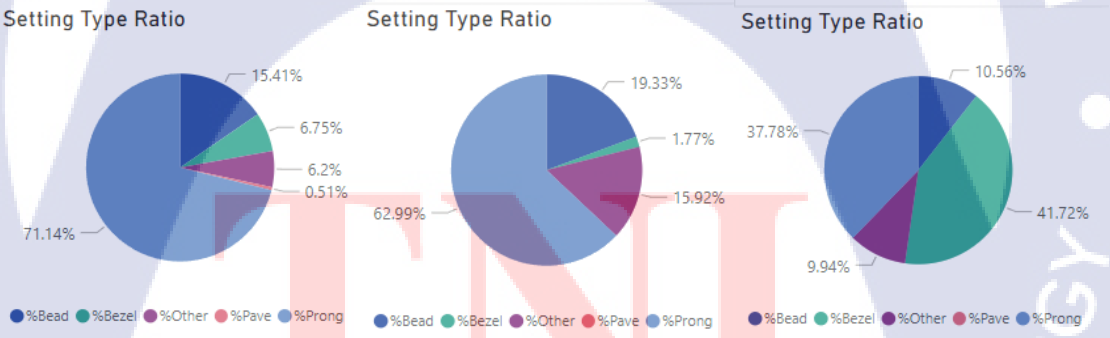


รูปที่ 0.5 ประสิทธิภาพการฝังอัญมณีของพนักงาน โดยรวมทุกประเภทการฝังในปี พ.ศ. 2564

3.3.3 ประเภทการฝัง

จากรูปที่ 3.6 ด้านซ้ายมือเป็นสัดส่วนการฝังอัญมณีแต่ละประเภทรวมทั้งแผนกในปี พ.ศ. 2564 พบว่าประเภทการฝังหนามเตยมีสัดส่วนมากที่สุด คือ 71.14% การฝังไข่ปลา และการฝังหุ้ม 15.41%, 6.75% ตามลำดับ โดยสัดส่วนของประเภทการฝังอัญมณีของทั้งแผนกนั้นมีผลมาจากรูปแบบของเครื่องประดับที่ถูกเปิดใบสั่งผลิตมา ซึ่งเมื่อมีชิ้นงานถูกส่งต่อมาจากกระบวนการก่อนหน้าเข้ามาที่แผนกฝังแล้ว พนักงานที่ทำหน้าที่แบ่งจ่ายงานให้กับพนักงานแต่ละคนโดยยึดความเชี่ยวชาญของพนักงานเป็นหลัก ทำให้พนักงานได้รับงานในลักษณะที่ไม่เหมือนกันดังรูปที่ 3.6 กลาง และขวา โดยแผนภาพวงกลมตรงกลางนั้นเป็นของพนักงานที่ได้อันดับที่ 5 จากการจัดอันดับในรูปที่ 3.5 ซึ่งพนักงานท่านนี้เป็นผู้ที่มีความเชี่ยวชาญในการฝังประเภทอื่นๆนอกจากประเภทหลักที่สนใจ และแผนภาพวงกลมด้านขวาเป็นของพนักงานที่ได้อันดับที่ 17 เป็นผู้ที่มีความเชี่ยวชาญในชิ้นงานที่มีการฝังหุ้มมากกว่าพนักงานท่านอื่น ดังนั้นจึงสามารถกล่าวได้ว่าประเภทการฝังนั้นมีผลต่อปริมาณงานที่ได้

และปัจจัยอื่นนอกเหนือจากที่สามารถเห็นได้จากแผนภาพสรุปข้อมูลกำลังการผลิตของแผนกฝังอัญมณี ในปี พ.ศ. 2564 ที่ผู้วิจัยได้จัดทำขึ้น จากการสอบถามพนักงานในแผนกฝังยังมีปัจจัยอื่นเข้ามาเกี่ยวข้องด้วย เช่น ประเภท และขนาดของอัญมณี ความแข็งของตัวเรือนซึ่งมีบันทึกอยู่ในการรับ-จ่ายชิ้นงาน รวมไปถึงความพร้อมของอุปกรณ์ ลักษณะโต๊ะทำงาน สภาพแวดล้อมในการทำงาน และความอคติของผู้จ่ายงานเข้ามาร่วมด้วย ซึ่งเป็นปัจจัยที่ไม่ถูกบันทึกไว้ในการรับ-จ่ายชิ้นงานบนระบบบริหารจัดการทรัพยากรภายในองค์กรของบริษัท



รูปที่ 0.6 สัดส่วนการฝังอัญมณีแต่ละประเภทรวมทั้งแผนก (ซ้าย),สัดส่วนการฝังอัญมณีแต่ละประเภทของพนักงานคนที่ 5 และ 17 ตามลำดับ (กลาง, ขวา)

3.4 รวบรวมข้อมูลที่ใช้ในการศึกษา

ผู้วิจัยรวบรวมข้อมูลที่ถูกบันทึกไว้ในระบบบริหารจัดการทรัพยากรภายในองค์กรของบริษัทย้อนหลังเป็นเวลา 5 ตั้งแต่ปี พ.ศ. 2560-2564 ข้อมูลที่รวบรวมมาประกอบด้วย 2 ส่วน คือ บันทึก

การรับ-จ่ายตัวเรือน และบันทึกการจ่ายอัญมณีให้กับพนักงานสำหรับแต่ละเบอร์งาน (Bag No.) ซึ่งข้อมูลอยู่ในรูปแบบไฟล์ .csv สามารถอธิบายรายละเอียดของข้อมูลตามตารางที่ 3.1 และ 3.2

Process	Location	Issue Date	Receipt Date	Bag	Findings	Kt	Clr	Wt Deduct	Issue-Qty	Issue-G Wt	Issue-Net Wt	Receive d-Qty	Receive d-G Wt	Received-Net Wt	Actual Loss Wt	% Actual Loss	Allowable Loss %	Loss for worker
6 Setting & Engrave	[redacted]	00/00/00	5/4/2017	91498	[redacted]		BGP		20	165.34	165.26	20	165.06	164.98	0.28	0.17	1	1.37
6 Setting & Engrave		00/00/00	25/11/2017	95129			BGP		15	82.13	82.07	15	81.96	81.9	0.17	0.2	1	0.65
6 Setting & Engrave		00/00/00	10/1/2017	90130		9	Y		1	8.14	7.75	1	7.95	7.56	0.19	2.49	1	-0.12
6 Setting & Engrave		00/00/00	10/1/2017	90253		9	Y		1	4.76	4.01	1	4.51	3.76	0.25	6.22	1	-0.21
6 Setting & Engrave		00/00/00	12/1/2017	88516		9	Y		6	6.78	5.8	6	6.53	5.55	0.25	4.31	1	-0.19
6 Setting & Engrave		00/00/00	12/1/2017	88560		9	Y		5	4.09	3.8	5	3.92	3.63	0.17	4.48	1	-0.13
6 Setting & Engrave		00/00/00	20/1/2017	88568		9	Y		5	5.96	5.55	5	5.79	5.38	0.17	3.06	1	-0.12
6 Setting & Engrave		00/00/00	20/1/2017	90382		9	Y		2	9.28	8.9	2	9.05	8.67	0.23	2.61	1	-0.15
6 Setting & Engrave		00/00/00	23/1/2017	90442		9	Y		10	14.73	14.69	10	14.56	14.52	0.17	1.15	1	-0.02
6 Setting & Engrave		00/00/00	23/1/2017	90546		9	Y		3	4.64	4.63	3	4.6	4.59	0.04	0.76	1	0.01
6 Setting & Engrave		00/00/00	23/1/2017	88977		9	Y		10	10.04	6.2	10	9.47	5.63	0.57	9.2	1	-0.51

รูปที่ 0.7 ตัวอย่างข้อมูลการรับ-จ่ายตัวเรือนของแผนกฝังอัญมณี

ตารางที่ 0.1 ความหมายของหัวข้อในตารางการรับ-จ่ายตัวเรือน

ชื่อคอลัมน์	ความหมาย
Process	ขั้นตอนการผลิต เช่น Filing, Pre-polishing, Setting
Location	ชื่อพนักงานที่รับ-จ่ายชิ้นงาน
Issue Date	วันที่จ่ายงาน
Receipt Date	วันที่รับชิ้นงานคืน
Bag	หมายเลขกำกับชิ้นงาน
Finding	ชื่อตัวเรือน
Kt	สูตรโลหะของชิ้นงาน
Clr	สีของชิ้นงาน
Wt Deduct	น้ำหนักที่คืน (กรณีที่คืนมาเป็นตัวเรือนเสีย)
Issue-Qty	จำนวนชิ้นที่จ่ายให้พนักงาน
Issue-Gwt	น้ำหนักชิ้นงานที่จ่าย (รวมน้ำหนักอัญมณี)
Issue-Nwt	น้ำหนักชิ้นงานที่จ่าย (คิดเฉพาะน้ำหนักโลหะ)
Received-Qty	จำนวนชิ้นงานที่รับคืนจากพนักงาน
Received-Gwt	น้ำหนักชิ้นงานที่รับคืน (รวมน้ำหนักอัญมณี)
Receive-Nwt	น้ำหนักชิ้นงานที่รับคืน (คิดเฉพาะน้ำหนักโลหะ)
Actual Loss wt	น้ำหนักโลหะที่สูญเสียที่เกิดขึ้นจริง
%Actual Loss	สัดส่วนน้ำหนักโลหะที่สูญเสียไป เทียบกับน้ำหนักโลหะที่ได้รับคืน
Allowable loss %	เปอร์เซ็นที่อนุญาตให้สูญเสีย
Loss for worker	น้ำหนักโลหะที่สูญเสีย หลังจากหักส่วนที่อนุญาตให้สูญเสีย

Setter	Date	Bag	Order	Code	Shp	Clr	Qual	Set-Qty	Set-Wt	Setting Kind	Cost Unit	Cost Total
[redacted]	9/1/2017	88515	10328	DIR-0.005	RND	H	F/P1	12	0.06	GRS	4	48
	9/1/2017	88515	10328	OPTPA-10X7	PEAR		A	12	16.03	PRL	9	108
	9/1/2017	89173	10411	ASSORTED				6	19.95	BZL	15	90
	9/1/2017	90223	10544	BO-GR164	FS	BL		2	0.75	BZL	8	16
	9/1/2017	90223	10544	CZR-2.5	RND	W	HS	1	0.12	SBZ	8	8
	10/1/2017	90130	10525	BOR-10X8(4)	REC	BL		1	1.92	GNT	15	15
	10/1/2017	90130	10525	CZR-1.75	RND	W	SIG	1	0.04	GRS	8	8
	10/1/2017	90251	10553	CZR-2.0	RND	W	SIG	1	0.06	GRS	3	3
	10/1/2017	90252	10553	CZR-1.25	RND	W	HS	4	0.07	GRS	3	12
	10/1/2017	90253	10554	CZR-5.5	RND	W	SIG	1	1.15	PRL	5	5
	11/1/2017	88348	10323	DIR-0.01	RND	H	F/P1	7	0.07	BZS	5	35
	11/1/2017	88348	10323	OPTOA-16X8	OVL		A	7	24.1	SBZ	15	105
	12/1/2017	88516	10328	OPTMA-8X4	MRQ		A	12	4.92	VPR	10	120
	12/1/2017	88560	10329	OPSPB-7X5	PEAR		S-B	5	1.47	VPR	9	45
	12/1/2017	90183	10534	OPTOC-8X6	OVL		C	10	8.74	BZL	15	150
	13/1/2017	90267	10562	DIR-0.005	RND	H	F/SI	16	0.06	GRS	5	80

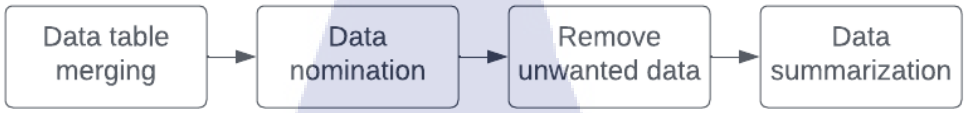
รูปที่ 0.8 ตัวอย่างข้อมูลการรับ-จ่ายอัญมณี

ตารางที่ 0.2 ความหมายของหัวข้อในตารางการรับ-จ่ายอัญมณี

ชื่อคอลัมน์	ความหมาย
Setter	ชื่อของพนักงานที่ถูกจ่ายงานให้
Date	วันที่จ่ายงาน
Bag	เบอร์งานในระบบ
Order	หมายเลขใบสั่งผลิต
Code	รหัสแทนชื่ออัญมณีที่จ่ายให้พนักงาน
Shp	รูปทรงของอัญมณีที่ใช้
Clr	สีของอัญมณี
Qual	คุณภาพของอัญมณี
Set-Qty	จำนวนอัญมณีที่จ่ายไป (เม็ด)
Set-Wt	น้ำหนักอัญมณีที่จ่ายไป (กะรัต)
Setting Kind	ประเภทการฝัง
Cost Unit	ราคาฝังต่อเม็ด (บาท)
Cost Total	ราคารวม (บาท)

3.5 การเตรียมข้อมูล

ในขั้นตอนกระบวนการเตรียมข้อมูลเป็นไปตามรูปที่ 3.8 ซึ่งสามารถอธิบายได้ดังนี้



รูปที่ 0.9 ขั้นตอนการเตรียมข้อมูล

3.5.1 รวมข้อมูลเป็นตารางเดียว (Data Table Merging)

โดยใช้ตารางบันทึกการรับ-จ่ายอัญมณีเป็นตารางหลัก เนื่องจากเป็นตารางที่มีข้อมูลที่สนใจอยู่มากกว่า ใช้คอลัมน์ [Bag No.] เป็นคีย์หลักในการดึงคอลัมน์ที่ต้องการจากบันทึกการรับ-จ่ายตัวเรือนเข้ามาไว้ในตารางเดียวกัน ซึ่งข้อมูลที่ต้องการคือ สูตรโลหะของชิ้นงาน [Kt] ต่อมาทำการจัดกลุ่มประเภทการฝังในหัวข้อ [Setting Kind] ประเภทที่มีความใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน เช่น BZS, BZM, BZL จะถูกจัดอยู่ในกลุ่มการฝังหุ้ม (Bezel setting) เนื่องจากใช้วิธีฝังหุ้มเหมือนกัน แตกต่างกันเพียงขนาดอัญมณีที่ใช้ในการฝังเท่านั้น

3.5.2 แปลงค่าข้อมูลที่สามารถจัดกลุ่ม (Data Nomination)

โดยแปลงข้อมูลตัวอักษรให้เป็นข้อมูลตัวเลขที่สามารถนำเข้าแบบจำลองวิเคราะห์ได้ เช่น ชื่อพนักงานฝัง [Setter], ประเภทการฝัง [Setting Kind], อัญมณีที่ใช้ในการฝัง [Code], ชนิดของโลหะตัวเรือน [Kt] ดังตารางที่ 3.3

3.5.3 กำจัดข้อมูลที่ไม่เกี่ยวข้องออก

โดยลบคอลัมน์ที่เป็นข้อมูลตัวอักษรออก ลบคอลัมน์ที่เป็นช่องว่าง และข้อมูลที่ไม่เกี่ยวข้องออก รวมไปถึงแถวที่มีค่าในคอลัมน์ [Set-Qty] ติดลบออก เนื่องจากแถวดังกล่าวเป็นบันทึกการคืนอัญมณีที่ไม่ได้ใช้ออกจากเบอร์งานในแถวนั้น

3.5.4 รวมบรรทัดที่มีรายละเอียดการทำงานที่เหมือนกันเข้าด้วยกัน

เช่น พนักงานฝัง, ประเภทการฝัง, อัญมณีที่ใช้ในการฝัง และชนิดของโลหะตัวเรือน เนื่องจากข้อมูลที่นำมาศึกษา เป็นข้อมูลบันทึกการรับ-จ่ายวัตถุดิบในการผลิตของแต่ละเบอร์งาน ดังนั้นข้อมูลที่ถูกบันทึกในแต่ละแถวจะยึดตามแต่ละเบอร์งาน ไม่ได้เป็นการสรุปของแต่ละวันมา เพื่อให้สะดวกต่อการนำข้อมูลไปใช้ในการศึกษา ซึ่งมีรายละเอียดตามตารางที่ 3.3

ตารางที่ 0.3 รายละเอียดข้อมูลหลังจากเตรียมข้อมูลเสร็จแล้ว

ชื่อคอลัมน์	คำอธิบาย	จำนวน Class	รายละเอียด Class
Date	วันที่รับตัวเรือนคืน	-	-
Metal	ชนิดของโลหะตัวเรือน	3	1 = เงิน 2 = ทอง 3 = แพลตินั่ม
SetKind	ประเภทการฝัง	5	1 = ฝังประเภทอื่นๆ 2 = ฝังไข่ปลา (Bead setting) 3 = ฝังหนามเตย (Prong setting) 4 = ฝังหุ้ม (Bezel setting) 5 = ฝังไข่ปลา (Pavé setting)
StoneKind	ประเภทอัญมณีที่ใช้	4	1 = อัญมณีสังเคราะห์ 2 = เพชร 3 = อัญมณีเนื้อแข็ง เช่น ทับทิม, แซฟไฟร์ 4 = อัญมณีเนื้ออ่อน เช่น มรกต, อเมทิสต์
Setter	พนักงานที่ฝังอัญมณี	19	1-18 ตามลำดับรหัสพนักงาน
Set-Qty	จำนวนอัญมณีที่ฝังได้ (เม็ด)	-	-

3.6 ทดลองหาแบบจำลองที่เหมาะสมกับการทำนายผลิตภาพแรงงาน

ผู้วิจัยแบ่งข้อมูลที่ผ่านการเตรียมข้อมูลมาตามหัวข้อที่ 3.5 เป็น 2 ส่วน ส่วนแรกเป็นข้อมูลระหว่างปี พ.ศ. 2560 – 2563 นำมาใช้ในการสร้างแบบจำลอง และนำข้อมูลส่วนที่ 2 มาเป็นชุดข้อมูลที่ใช้ในการทดสอบการทำงานของแบบทดสอบ ซึ่งเป็นข้อมูลในปี พ.ศ. 2564 ข้อมูลแต่ละชุดมีจำนวนปีของข้อมูลต่างกันชุดละ 1 ปี โดยให้ความสำคัญกับข้อมูลที่มีความใกล้เคียงปีปัจจุบันมากกว่า ชุดข้อมูลที่ได้จะมีรายละเอียดตามตารางที่ 3.4

ชุดข้อมูลที่ 1-4 แต่ละชุดจะถูกแบ่งเป็น 2 ส่วนเพื่อนำไปใช้สอนและทดสอบแบบจำลอง (Train-Test Split) ด้วยอัตราส่วน 80:20 ตามลำดับ ต่อมาทำการสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เทคนิคป่าสุ่ม (Random Forest) และเทคนิคเกรเดียนท์บูสติ้ง (Gradient Boosting) ด้วยภาษาโปรแกรมไพทอน

(Python programming language) ทุกเทคนิคจะถูกวัดประสิทธิภาพด้วยค่า Adjust R-Squared และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) หลังจากสร้างแบบจำลองเสร็จเรียบร้อยแล้ว ทำการเปรียบเทียบผลของแต่ละแบบจำลองที่ได้ แล้วเลือกเทคนิคที่มีประสิทธิภาพสูงสุดจากแบบจำลองที่มีค่า Adjust R-Squared สูงที่สุด และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ยต่ำที่สุดไปทดสอบการทำนายโดยการนำชุดข้อมูลที่ 5 มาเป็นชุดข้อมูลทดสอบ

ตารางที่ 0.4 ตารางแสดงรายละเอียดของชุดข้อมูลที่ถูกแบ่งไปใช้สร้างแบบจำลอง และทดสอบ

ชุดข้อมูล	**รายละเอียดชุดข้อมูล**	**ขนาดของข้อมูล (แถว, คอลัมน์)**
ชุดข้อมูลที่ 1	ข้อมูลระหว่างปี 2560-2563	(11815, 9)
ชุดข้อมูลที่ 2	ข้อมูลระหว่างปี 2561-2563	(10490, 9)
ชุดข้อมูลที่ 3	ข้อมูลระหว่างปี 2562-2563	(9222, 9)
ชุดข้อมูลที่ 4	ข้อมูลปี 2563	(4714, 9)
ชุดข้อมูลที่ 5	ข้อมูลปี 2564	(5935, 9)

3.7 สรุปผลและจัดทำวิทยานิพนธ์

ผู้วิจัยเปรียบเทียบแบบจำลองที่นำมาใช้ในการทำนายผลิตภาพแรงงานของแผนกฝังอัญมณี และสรุปผลว่าแบบจำลองใดมีประสิทธิภาพมากที่สุด และมีความเหมาะสมกับข้อมูลที่มีอยู่ รวมไปถึงสรุปถึงปัจจัยที่มีผลต่อค่าผลิตภาพแรงงานในขั้นตอนการฝังอัญมณีตามวัตถุประสงค์ในบทที่ 1

บทที่ 4
ผลการวิจัย

การศึกษาวิจัยการทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยการใช้การเรียนรู้ของเครื่อง ได้ทำการศึกษาเปรียบเทียบดังนี้

4.1 ผลการเปรียบเทียบแบบจำลองการทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้การเรียนรู้ของเครื่อง

จากการนำเทคนิคการเรียนรู้ของเครื่องจำนวน 4 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เทคนิคป่าสุ่ม (Random Forest) และเทคนิคเกรเดียนท์บูสติ้ง (Gradient Boosting) ได้ผลการศึกษาดังนี้

4.1.1 <u>ส่วนที่ 1 ศึกษาเปรียบเทียบแบบจำลองจากการสร้างแบบจำลอง 4 ครั้ง ทำการประเมินประสิทธิภาพของแบบจำลองโดยใช้ค่า Adjusted R-Squared และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error: RMSE)</u>

จากการทดลองจำนวน 4 ครั้ง เพื่อทดสอบว่าแบบจำลองใดมีประสิทธิภาพมากที่สุด สามารถสรุปผลการทดลองได้ดังตารางที่ 4.1 และ 4.2

การทดลองครั้งที่ 1 ด้วยชุดข้อมูลของปี พ.ศ. 2560-2563 ที่มีขนาด 11,815 แถว 9 คอลัมน์ พบว่าเทคนิคป่าสุ่ม (Random Forest) มีประสิทธิภาพในการทำนายค่าได้ดีที่สุด ได้ค่า Adjusted R-Squared เป็น 0.5880 และค่า RMSE = 0.1874

การทดลองครั้งที่ 2 ด้วยชุดข้อมูลของปี พ.ศ. 2561-2563 ที่มีขนาด 10,490 แถว 9 คอลัมน์ พบว่าเทคนิคป่าสุ่ม (Random Forest) มีประสิทธิภาพในการทำนายค่าได้ดีที่สุด ได้ค่า Adjusted R-Squared เป็น 0.5784 และค่า RMSE = 0.1884

การทดลองครั้งที่ 3 ด้วยชุดข้อมูลของปี พ.ศ. 2562-2563 ที่มีขนาด 9,222 แถว 9 คอลัมน์ พบว่าเทคนิคป่าสุ่ม (Random Forest) มีประสิทธิภาพในการทำนายค่าได้ดีที่สุด ได้ค่า Adjusted R-Squared เป็น 0.5333 และค่า RMSE = 0.1937

การทดลองครั้งที่ 4 ด้วยชุดข้อมูลของปี พ.ศ. 2563 ที่มีขนาด 4,714 แถว 9 คอลัมน์ พบว่าเทคนิคป่าสุ่ม (Random Forest) มีประสิทธิภาพในการทำนายค่าได้ดีที่สุด ได้ค่า Adjusted R-Squared เป็น 0.4250 และค่า RMSE = 0.1967

ตารางที่ 0.1 ผลการเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณีด้วยค่า Adjusted R-Square

แบบจำลอง	**ชุดข้อมูลที่ 1 (2560-2563)**	**ชุดข้อมูลที่ 2 (2561-2563)**	**ชุดข้อมูลที่ 3 (2562-2563)**	**ชุดข้อมูลที่ 4 (2563)**
Decision Tree	0.4531	0.4290	0.3467	0.2290
K-Nearest Neighbors	0.4804	0.4449	0.3947	0.3187
Random Forest	**0.5880**	**0.5784**	**0.5333**	**0.4250**
Gradient Boosting	0.5590	0.5050	0.4760	0.3240

ตารางที่ 0.2 ผลการเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณีด้วยค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย

แบบจำลอง	**ชุดข้อมูลที่ 1 (2560-2563)**	**ชุดข้อมูลที่ 2 (2561-2563)**	**ชุดข้อมูลที่ 3 (2562-2563)**	**ชุดข้อมูลที่ 4 (2563)**
Decision Tree	20.7690	21.5485	28.1022	29.8466
K-Nearest Neighbors	21.7913	22.0419	28.6123	32.4398
Random Forest	**0.1874**	**0.1884**	**0.1937**	**0.1967**
Gradient Boosting	20.4991	20.8430	26.6380	30.0566

จากการสร้างแบบจำลองทั้ง 4 ครั้ง พบว่าเทคนิคป่าสุ่ม (Random Forest) มีประสิทธิภาพในการทำนายสูงที่สุด มีความน่าจะเป็นที่จะสามารถทำนายค่าได้แม่นยำที่สุดเมื่อมีการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตที่มีจำนวนมาก เนื่องจากเทคนิคดังกล่าวเหมาะกับชุดข้อมูลที่สามารถมีคำตอบได้ในช่วงกว้าง หรือมีคำตอบที่หลากหลาย

4.1.2 <u>ส่วนที่ 2 ทดลองใช้แบบจำลองที่มีประสิทธิภาพสูงสุดกับข้อมูลชุดที่ 5</u>

จากการทดลองนำแบบจำลองมาทดลองใช้กับข้อมูลชุดที่ 5 ที่มีขนาด 5,935 แถว 9 คอลัมน์ พบว่าแบบจำลองได้ค่า Adjusted R-Squared = 0.5290 และค่า RMSE = 0.2113 ซึ่งมีความใกล้เคียงกับชุดข้อมูลที่ 1-4 ที่ใช้เทคนิคป่าสุ่ม (Random Forest) เหมือนกัน

4.2 ปัจจัยที่มีผลต่อการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณี

จากการศึกษาปัจจัยที่มีความสำคัญต่อการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณี ด้วยเทคนิค Feature Importance โดยปัจจัยที่มีน้ำหนักมากคือปัจจัยที่มีประโยชน์ และมีผลต่อประสิทธิภาพในการทำนายข้อมูล ส่วนปัจจัยที่มีน้ำหนักน้อยเป็นปัจจัยที่มีผลต่อการทำนายน้อย

ตารางที่ 0.3 ปัจจัยที่มีผลต่อการทำนายค่าผลิตภาพของแผนกฝังอัญมณี

ลำดับที่	ปัจจัย	ค่าน้ำหนัก
1	พนักงานฝัง	0.2983
2	เดือน	0.2114
3	วันในแต่ละสัปดาห์	0.1499
4	ประเภทการฝัง	0.1392
5	ประเภทอัญมณี	0.0907
6	ปี	0.0828
7	โลหะของตัวเรือน	0.0275

จากตารางที่ 4.3 พบว่าปัจจัยที่มีผลมากที่สุด 4 อันดับแรก คือ พนักงานฝัง, เดือน, วันในแต่ละสัปดาห์, ประเภทการฝัง และได้ค่าความสัมพันธ์เท่ากับ 0.2983, 0.2114, 0.1499 และ 0.1392 ตามลำดับ เนื่องมาจากสาเหตุต่อไปนี้

อันดับที่ 1 พนักงาน : เนื่องจากกระบวนการฝังอัญมณีเป็นกระบวนการผลิตที่เกิดจากฝีมือของมนุษย์ ดังนั้นจึงมีเรื่องความสามารถ มีทักษะส่วนตัว รวมไปถึงประสบการณ์ที่แตกต่างกันออกไปของพนักงานแต่ละคนเข้ามาเกี่ยวข้อง

อันดับที่ 2 ช่วงเดือนในแต่ละปี : เนื่องจากกลยุทธ์ข้อหนึ่งของบริษัทคือ การผลิตเมื่อคำสั่งผลิต (Made to order) จึงมีแนวโน้มที่จะมียอดสั่งผลิตเพิ่มมากขึ้นในช่วงก่อนใกล้ถึงเทศกาลสำคัญต่างๆ เพื่อใช้เครื่องประดับอัญมณีเป็นของขวัญตามเทศกาล ดังนั้นเมื่อมียอดการผลิตเพิ่มขึ้น ก็ทำให้ปริมาณการฝังเพิ่มมากขึ้นได้ ซึ่งแตกต่างกับบริษัทที่ผลิตเพื่อเก็บเข้าคลังสินค้า (Made to stock)

อันดับที่ 3 วันในแต่ละสัปดาห์ : เนื่องจากการกำหนดรอบการส่งออกชิ้นงานในทุกวันศุกร์ของสัปดาห์ ทำให้ช่วงต้นสัปดาห์มักปริมาณการฝังไม่สูง เพราะปริมาณงานที่ตกค้างจากสัปดาห์ก่อนมีไม่มาก และจะค่อยๆเพิ่มขึ้นจนกระทั่งวันพฤหัสบดีที่มีปริมาณการฝังสูงที่สุดในสัปดาห์ และมีการลดลงในวันศุกร์ และวันเสาร์ที่มีการทำล่วงเวลา

อันดับที่ 4 ประเภทการฝัง : เนื่องจากวิธีการฝังแต่ละประเภทมีรายละเอียดที่แตกต่างกัน เช่น วิธีการเตรียมชิ้นงานก่อนฝัง, ขนาดของพื้นที่ผิวชิ้นงานที่พนักงานต้องจัดการ เป็นต้น

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษาวิจัยเรื่องการทำนายผลิตภาพแรงงานในกระบวนการฝังอัญมณีโดยใช้เทคนิคการเรียนรู้ของเครื่องซึ่งมีวัตถุประสงค์ตามบทที่ 1 สามารถสรุปได้ตามหัวข้อต่อไปนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ปัญหาและอุปสรรคในการวิจัย

5.4 ข้อเสนอแนะงานวิจัย

5.1 สรุปผลการวิจัย

5.1.1 ผลการศึกษาเปรียบเทียบแบบจำลองการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณีโดยใช้การเรียนรู้ของเครื่องด้วย 4 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เทคนิคป่าสุ่ม (Random Forest) และเทคนิคเกรเดียนท์บูสติ้ง (Gradient Boosting) สามารถสรุปผลการศึกษาได้ดังนี้

(1) การเปรียบเทียบแบบจำลอง จากการสร้างแบบจำลอง 4 ครั้ง ด้วยเทคนิคการเรียนรู้ของเครื่องจำนวน 4 เทคนิค พบว่า การสร้างแบบจำลองในครั้งที่ 1 ที่ใช้ชุดข้อมูลในปี 2560-2563 นั้นให้ค่า Adjusted R-Square สูงที่สุด ค่า RMSE ต่ำที่สุด และเมื่อตัดข้อมูลปีเก่าออกครั้งละ 1 ปี ทำให้ค่า Adjusted R-Square ลดลง และค่า RMSE สูงขึ้น จากการทดลองนี้สรุปได้ว่า ข้อมูลในอดีตที่นำมาใช้ในการทดสอบสอบมีผลต่อการทำนายค่า

(2) การเปรียบเทียบแบบจำลอง จากการสร้างแบบจำลอง 4 ครั้ง ด้วยชุดข้อมูลจำนวน 4 ชุด พบว่าเทคนิคป่าสุ่ม (Random Forest) ให้ค่า Adjusted R-Square สูงที่สุด ค่า RMSE ต่ำที่สุด เมื่อเปรียบเทียบกับการใช้เทคนิคอื่นๆ เนื่องจากเทคนิคดังกล่าวเหมาะกับชุดข้อมูลที่สามารถมีคำตอบได้ในช่วงกว้าง หรือมีคำตอบที่หลากหลาย ทำให้เทคนิคป่าสุ่ม (Random Forest) เป็นเทคนิคที่เหมาะสมในการนำมาประยุกต์ใช้กับลักษณะของข้อมูลที่มีอยู่

5.1.2 การศึกษาปัจจัยที่มีผลต่อการทำนายค่าผลิตภาพแรงงานของแผนกฝังอัญมณี พบว่าปัจจัยที่มีผลมากที่สุดคือ ความสามารถของพนักงานฝัง รองลงมาเป็นช่วงเวลาที่งานเข้ามาในหน่วยงานฝังในหน่วย เดือน และวัน ซึ่งสามารถกล่าวได้ว่าเป็นรอบของการผลิตในแต่ละบริษัท รองลงมาเป็นประเภทของการฝัง และอัญมณีที่ใช้ในการฝัง และชนิดของโลหะตัวเรือนเป็นปัจจัยที่มีผลน้อยที่สุด

5.2 อภิปรายผลการวิจัย

5.2.1 ขั้นตอนการทำความสะอาดและเตรียมข้อมูล

(1) จากข้อมูลที่ถูกบันทึกไว้ในระบบบริหารจัดการทรัพยากรภายในองค์กรของบริษัทที่ผู้วิจัยนำมาใช้ในการศึกษานั้น ในกระบวนการทำความสะอาดข้อมูล พบว่าในตารางบันทึกการรับ-จ่ายอัญมณีนั้นเป็นการบันทึกรวมทั้งการจ่ายอัญมณีให้พนักงาน และการรับคืนจากพนักงาน ในกรณีที่มีการจ่ายอัญมณีเกินจำนวน จ่ายให้ผิดขนาด หรือผิดจากลักษณะที่ถูกกำหนดมา รวมไปถึงการคืน อัญมณีเนื่องจากพนักงานทำอัญมณีแตก โดยในชุดข้อมูลไม่มีได้มีการระบุว่าเป็นการจ่ายให้พนักงาน หรือรับคืนจากพนักงานอย่างชัดเจน แต่มีความแตกต่างกันที่ ถ้าเป็นการจ่ายให้พนักงาน คอลัมน์ [Set-Qty] (จำนวนอัญมณี) และ [Set-Wt] (น้ำหนักอัญมณี) จะมีค่าเป็นจำนวนเต็มบวก แต่ถ้าเป็นการรับคืนจากพนักงานจะเป็นจำนวนเต็มลบแทน ดังนั้น ในการนำข้อมูลมาใช้ต้องทำความเข้าใจข้อมูลที่มี และสังเกตความแตกต่างระหว่างการรับ-จ่ายให้ได้ เพื่อให้สามารถตัดข้อมูลที่ไม่เกี่ยวข้องกับการวิจัยออกได้อย่างถูกต้อง และสามารถนำข้อมูลไปใช้ได้อย่างมีประสิทธิภาพ

(2) จากข้อมูลที่นำมาใช้พบว่าคอลัมน์ [SetKind] (ประเภทการฝัง) นั้นมีการใช้ชื่อเรียกหลากหลาย ขึ้นอยู่กับผู้ที่สร้างรายละเอียดงานเข้าไปในระบบบริหารจัดการทรัพยากรภายในองค์กรของบริษัท ซึ่งจากข้อมูลที่นำมาใช้ศึกษาพบว่ามีมากกว่า 70 ชื่อที่ไม่ซ้ำกัน โดยบางชื่อก็เป็นวิธีการฝังแบบเดียวกัน แต่เรียกชื่อต่างกันเพราะเป็นชื่อของคนละลูกค้ากัน มีการตั้งค่าราคาต้นทุนไว้ไม่เท่ากัน และสาเหตุอื่นๆที่ทำให้มีชื่อเรียกวิธีการฝังเดียวกันอย่างหลากหลาย ทำให้ยากต่อการเข้าใจและจัดกลุ่มข้อมูลในคอลัมน์ ดังนั้นในกระบวนการจัดกลุ่มข้อมูลของคอลัมน์ประเภทการฝังจึงต้องทำความเข้าใจถึงสาเหตุและความแตกต่างของการฝังแต่ละประเภท จึงจะสามารถจัดกลุ่มข้อมูลได้อย่างถูกต้อง รวมไปถึงในการออกแบบชื่อประเภทการฝังอัญมณีในระบบบริหารจัดการทรัพยากรภายในองค์กร ถ้าเป็นวิธีการฝังเดียวกันควรใช้ชื่อร่วมกัน เพื่อป้องกันการสับสนของผู้ที่มาเริ่มใช้งานระบบในภายหลัง รวมไปถึงจะง่ายต่อการทำความเข้าใจ และนำข้อมูลไปใช้ในการวิเคราะห์ต่อไปด้วย

5.2.2 ประโยชน์ที่ได้จากงานวิจัย

(1) แผนภาพสรุปข้อมูลกำลังการผลิตของแผนกฝังอัญมณี ตามรูปที่ 3.2 ในบทที่ 3 สามารถนำไปใช้เป็นต้นแบบในการสร้างแผนภาพของหน่วยงานผลิตอื่นๆต่อไปได้ และเมื่อมองในอีกมุมมอง อาจทำให้เห็นถึงโอกาสในการรับงานฝังอัญมณีเพิ่มเข้ามาในช่วงที่มีคำสั่งผลิตจากลูกค้าเข้ามาน้อยได้ รวมไปถึงจากรูปที่ 3.5 ที่ได้แสดงประสิทธิภาพการฝังอัญมณีของพนักงานไว้ ช่วยให้สามารถเข้าใจถึงความสามารถโดยเบื้องต้นของพนักงาน และดำเนินแผนการพัฒนาความสามารถของพนักงานรายบุคคลได้

(2) ผลที่ได้จากการทำนายค่าผลิตภาพแรงงานสามารถนำไปใช้เป็นข้อมูลประกอบการตัดสินใจในการทำงานได้ เช่น ฝ่ายวางแผนนำไปใช้ในการวางแผนการผลิตล่วงหน้า, ฝ่ายขายสามารถประเมินระยะเวลาในการผลิตเบื้องต้นให้กับลูกค้า, ฝ่ายวิจัยและพัฒนาสามารถนำไปใช้เป็นข้อมูลตั้งต้น หรือดัชนีชี้วัดในการทำกิจกรรมพัฒนา ปรับปรุงกระบวนการ รวมไปถึงการนำไปใช้ในการคำนวณต้นทุนในการผลิตชิ้นงาน เป็นต้น

(3) จากตารางที่ 4.3 แสดงให้เห็นถึงปัจจัยที่มีผลในการเพิ่ม-ลดของค่าผลิตภาพแรงงานได้อย่างชัดเจนมากยิ่งขึ้น ช่วยให้สามารถตัดสินใจเลือกดำเนินการพัฒนาเฉพาะจุดได้ดียิ่งขึ้น

5.3 ปัญหาและอุปสรรคในการทำวิจัย

ผู้วิจัยต้องศึกษาการใช้ภาษาโปรแกรมไพทอนเพิ่มเติมเพื่อนำมาสร้างแบบจำลองจากแต่ละเทคนิคการเรียนรู้ของเครื่อง ซึ่งเป็นภาษาที่ผู้วิจัยยังมีประสบการณ์ไม่มากนัก

5.4 ข้อเสนอแนะงานวิจัย

5.4.1 จากการศึกษาพบว่าปัจจัยที่มีผลต่อการทำนายมากที่สุด 3 อันดับแรกจากตารางที่ 4.3 คือ ความแตกต่างของพนักงานฝัง, ช่วงเวลาในหน่วยเดือน วัน และประเภทของการฝัง ดังนั้นในการเตรียมข้อมูลจึงควรให้ความสนใจกับปัจจัยดังกล่าว และเมื่อนำแบบจำลองไปใช้งานจึงไม่จำเป็นต้องให้ความสนใจกับปัจจัยอื่นมากเท่ากับปัจจัย 3 อันดับแรก เพื่อที่จะได้ไม่ต้องเสียเวลาในการจัดการกับปัจจัยที่ไม่มีผลต่อประสิทธิภาพของการทำนายข้อมูล

5.4.2 เนื่องจากข้อมูลที่นำมาศึกษาเป็นข้อมูลจากทางฝ่ายผลิตเพียงส่วนเดียว ทำให้ไม่มีข้อมูลในด้านของทักษะ ความสามารถ ประสบการณ์ของพนักงานฝังมาใช้ในการวิเคราะห์ร่วมด้วย ดังนั้นถ้ามีข้อมูลดังกล่าวเข้ามาเพิ่มในการศึกษาอาจทำให้สามารถเพิ่มประสิทธิภาพในการทำนายได้อีกด้วย

5.4.3 นำไปแบบจำลองประยุกต์ใช้กับฝ่ายผลิตในกระบวนการอื่นๆ เพื่อใช้ในการปรับปรุงกระบวนการผลิตต่อไป

บรรณานุกรม

บรรณานุกรม

[1] สำนักงานปลัดกระทรวงพาณิชย์, “*สินค้าส่งออกสำคัญของไทยรายประเทศ,*” [Online]. Available: https://tradereport.moc.go.th/Tradethai.aspx. [Accessed: 15 มีนาคม 2565].

[2] A. A. Skuratowicz et al., “Judging the quality of stone setting,” in *The SantaFe Symposium*, New Mexico, USA, May 2004, pp. 435-450.

[3] S. Thaowongsa, “Effect of basic gems-setting types on jewelry design,” *Journal of the Faculty of Architecture King Mongkut's Institute of Technology Ladkrabang,* vol. 18, no. 1, pp. 91-104, June 2014.

[4] A. Unwin, “Why is data visualization important? what is important in data visualization?,” *Harvard Data Science Review,* vol. 1, no. 2, pp. 1-7, January 2020.

[5] H. Kaur, “Why Data Visualization Matters in Data Analytics?,” [Online]. Available: https://www.geeksforgeeks.org/why-data-visualization-matters-in-data-analytics/. [Accessed: February 13, 2022].

[6] J. Quinlan, “Induction of decision trees,” *Machine Learning,* vol. 1, no. 1, pp. 81-106, March 1986.

[7] J. Korstanje, *Advanced Forecasting with Python*, Maisons Alfort, France: APress, 2021.

[8] C. Patcharacharoenwong et al., “Arrival time prediction model to a pier for public transportation boats,” *Journal of Science Ladkrabang,* vol. 29, no. 2, pp. 31-44, December 2020.

[9] P. Teerarassamee, *The Methodology to Find Appropriate K for K-Nearest Neighbor Classification with Medical Datasets*, Nakhon Ratchasima: Suranaree University of Technology, 2015.

[10] S. Ebrahimi et al., “Hybrid artificial intelligence HFS-RF-PSO model for construction labor productivity prediction and optimization,” *Algorithms,* vol. 14, no. 214, pp. 1-18, July 2021.

[11] F. Mirahadi and T. Zayed, “Simulation-based construction productivity forecast using Neural-Network-Driven Fuzzy Reasoning,” *Automation in Construction,* vol. 65, no. 11, pp. 102-115, May 2016.

[12] S. Supsomboon, “Simulation for jewelry production process improvement using line balancing: A case study,” *Management Systems in Production Engineering,* vol. 27, no. 3, pp. 127-137, August 2019.

[13] S. Ray and M. Abotaleb, “Modeling and forecasting of producer price index (PPI) of cheese manufacturing industry,” *Journal of Agriculture, Biology and Applied Statistics,* vol. 1, no. 1, pp. 39-49, July 2022.

[14] M. Oral et al., “Supervised vs. unsupervised learning for construction crew productivity prediction,” *Automation in Construction,* vol. 22, no. 25, pp. 271-276, March 2022.

[15] E. Elwakil and T. Zayed, “An innovative technique for improving productivity forecasting models,” in *4th Construction Specialty Conference*, Montréal, Québec, CSCE 2013, May 29 - June 1, 2013, pp. 1-10.

[16] Z. Dou et al., “Regional manufacturing industry demand forecasting: A deep learning approach,” *Applied Science,* vol. 11, no. 13, pp. 1-15, July 2021.

[17] P. DiCristofaro, “The history of American ring making, 1798-2018,” in *The SantaFe Symposium*, New Mexico, USA, May 2019, pp. 203-238.

[18] S. W. Kielarova et al., “Development of computer-aided design module for automatic gemstone setting on halo ring,” *KKU Engineering Journal,* vol. 43, no. S2, pp. 239-243, June 2016.

[19] J. Taylor, “Designing bench-friendly models,” in *The SantaFe Symposium*, New Mexico, USA, May 2017, pp. 441-472.

[20] C. Corti, “The role of hardness in jewelry alloys,” in *The SantaFe Symposium*, New Mexico, USA, May 2008, pp. 103-120.

[21] B. Lohwongwatana and E. Nisaratanaporn, “On hardness,” in *The SantaFe Symposium*, New Mexico, USA, May 2010, pp. 351-362.

[22] N. L. Attaway and S. W. Attaway, “Back from the dead: how to resurrect dead gemstones,” in *The SantaFe Symposium*, New Mexico, USA, May 2018, pp. 33-70.

[23] R. S. Priya and V. Aroulmoji, “A review on productivity and its effect in industrial manufacturing,” *International Journal of Advanced Science and Engineering,* vol. 6, no. 4, pp. 1490-1499, May 2020.

[24] M. Sreekumar et al., “Productivity in manufacturing industries,” *International Journal of Innovative Science and Research Technology,* vol. 3, no. 10, pp. 634-639, October 2018.

ภาคผนวก

ภาคผนวก ก.
ชุดคำสั่งภาษาโปรแกรมไพทอน

```
#DecisionTree
#import data
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('train๑.csv') #filename

ax = data['Set-Qty'].plot()
ax.set_ylabel('Qty (Pcs)')
ax.set_xlabel('Time')
plt.show()

data.head()

#remove outliers
import numpy as np
Q๑ = np.percentile(data['Set-Qty'], ๒๕,
                   interpolation = 'midpoint')
Q๓ = np.percentile(data['Set-Qty'], ๗๕,
                   interpolation = 'midpoint')
IQR = Q๓ - Q๑
print("Old Shape: ", data.shape)
# Upper bound
upper = np.where(data['Set-Qty'] >= (Q๓+๑.๕*IQR))
# Lower bound
lower = np.where(data['Set-Qty'] <= (Q๑-๑.๕*IQR))
''' Removing the Outliers '''
data.drop(upper[๐], inplace = True)
data.drop(lower[๐], inplace = True)
print("New Shape: ", data.shape)

data['SQ๑'] = data['Set-Qty'].shift(๑)
data['SQ๒'] = data['Set-Qty'].shift(๒)
data['SQ๓'] = data['Set-Qty'].shift(๓)
data['SQ๔'] = data['Set-Qty'].shift(๔)
data['STK๑'] = data['StoneKind'].shift(๑)
data['STK๒'] = data['StoneKind'].shift(๒)
```

```
data['STK๓'] = data['StoneKind'].shift(๓)
data['STK๔'] = data['StoneKind'].shift(๔)
data['SEK๑'] = data['SetKind'].shift(๑)
data['SEK๒'] = data['SetKind'].shift(๒)
data['SEK๓'] = data['SetKind'].shift(๓)
data['SEK๔'] = data['SetKind'].shift(๔)
data['M๑'] = data['Metal'].shift(๑)
data['M๒'] = data['Metal'].shift(๒)
data['M๓'] = data['Metal'].shift(๓)
data['M๔'] = data['Metal'].shift(๔)

data = data.dropna()
X = data[['Year', 'Month', 'WeekDay','Setter', 'Metal', 'SetKind', 'StoneKind']]
y = data['Set-Qty']

# apply a min max scaler
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data = pd.DataFrame(scaler.fit_transform(data), columns = data.columns)

#Fitting the model
# Create Train test split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=๐.๒๐,
random_state=๑, shuffle=True)
from sklearn.tree import DecisionTreeRegressor
my_dt = DecisionTreeRegressor(random_state=๑๒๓๔๕)
my_dt.fit(X_train, y_train)
dt_fcst = my_dt.predict(X_test)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(dt_fcst, X_test๒).fit()
#est๒ = est.fit()
print("Adj. R๒ :\n",est.rsquared_adj)
```

```
#Adding a grid search
from sklearn.model_selection import GridSearchCV
my_dt = GridSearchCV(DecisionTreeRegressor(random_state=๔๔),
{'min_samples_split': list(range(๒๐,๕๐, ๒)),
'max_features': [๐.๖, ๐.๗, ๐.๘, ๐.๙, ๑.],
'criterion': ['mse', 'mae']},
scoring = 'r๒', n_jobs = -๑)
my_dt.fit(X_train, y_train)
dt_fcst = my_dt.predict(X_test)
#print(r๒_score(list(y_test), list(my_dt.predict(X_test))))

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(dt_fcst, X_test๒).fit()
#est๒ = est.fit()
adj_r๒ = est.summary()
print("Adj. R๒ :\n",adj_r๒)


#Finding the best parameters
print(my_dt.best_estimator_)

#RMSE
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(y_test, dt_fcst))
print("RMSE :\n", rmse)

# Listing ๑๒-๖. Plotting the prediction
dt_fcst = my_dt.predict(X_test)
plt.figure(figsize=(๒๐,๕))
plt.plot(list(y_test))
plt.plot(list(dt_fcst))
plt.ylabel('Sales')
```

```
plt.xlabel('Time')
plt.show()

#RandomForest
# Listing ๑๔-๑. Importing the data
import pandas as pd
data = pd.read_csv('train๑.csv') #filename

#remove outliers
import numpy as np
Q๑ = np.percentile(data['Set-Qty'], ๒๕,
                   interpolation = 'midpoint')
Q๓ = np.percentile(data['Set-Qty'], ๗๕,
                   interpolation = 'midpoint')
IQR = Q๓ - Q๑
print("Old Shape: ", data.shape)
# Upper bound
upper = np.where(data['Set-Qty'] >= (Q๓+๑.๕*IQR))
# Lower bound
lower = np.where(data['Set-Qty'] <= (Q๑-๑.๕*IQR))
''' Removing the Outliers '''
data.drop(upper[๐], inplace = True)
data.drop(lower[๐], inplace = True)
print("New Shape: ", data.shape)

# Adding a year of lagged data
data['L๑'] = data['Set-Qty'].shift(๑)
data['L๒'] = data['Set-Qty'].shift(๒)
data['L๓'] = data['Set-Qty'].shift(๓)
data['L๔'] = data['Set-Qty'].shift(๔)
data['L๕'] = data['Set-Qty'].shift(๕)
data['L๖'] = data['Set-Qty'].shift(๖)
data['L๗'] = data['Set-Qty'].shift(๗)
data['L๘'] = data['Set-Qty'].shift(๘)
data['L๙'] = data['Set-Qty'].shift(๙)
data['L๑๐'] = data['Set-Qty'].shift(๑๐)
```

```
data['L๑๑'] = data['Set-Qty'].shift(๑๑)
data['L๑๒'] = data['Set-Qty'].shift(๑๒)
data.head()

# apply a min max scaler
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data = pd.DataFrame(scaler.fit_transform(data), columns = data.columns)



# Fitting the default Random Forest regressor
# Create X and y object
data = data.dropna()
y = data['Set-Qty']
#X = data[['Year', 'Month', 'WeekDay', 'Setter', 'Metal', 'SetKind', 'StoneKind']]
X = data[['Month', 'WeekDay', 'Setter', 'SetKind', 'L๑', 'L๒', 'L๓', 'L๔', 'L๕', 'L๖']]
# , 'L๗', 'L๘', 'L๙', 'L๑๐', 'L๑๑', 'L๑๒']]

# Create Train test split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=๐.๒,
random_state=๑๒๓๔๕, shuffle=True)
from sklearn.ensemble import RandomForestRegressor
my_rf = RandomForestRegressor()
my_rf.fit(X_train, y_train)
rf_fcst = my_rf.predict(X_test)

#R๒
from sklearn.metrics import r๒_score
r๒ = r๒_score(list(y_test), list(rf_fcst))
print(r๒)
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(rf_fcst, X_test๒).fit()
```

```
#est๒ = est.fit()
print("Adj. R๒ :\n",est.rsquared_adj)

# Fitting the Random Forest regressor with hyperparameter tuning
from sklearn.model_selection import GridSearchCV
my_rf = GridSearchCV(RandomForestRegressor(), {'max_features':[๐.๖๕, ๐.๗, ๐.๗๕,
๐.๘, ๐.๘๕, ๐.๙, ๐.๙๕], 'n_estimators': [๑๐, ๕๐, ๑๐๐, ๒๕๐, ๕๐๐, ๗๕๐, ๑๐๐๐]}, scoring
= 'r๒', n_jobs = -๑)
my_rf.fit(X_train, y_train)
rf_fcst = my_rf.predict(X_test)
print(my_rf.best_params_)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(rf_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Obtaining the plot of the forecast on the test data
import matplotlib.pyplot as plt
plt.figure(figsize=(๒๐,๕))
plt.plot(list(rf_fcst))
plt.plot(list(y_test))
plt.legend(['fcst', 'actu'])
plt.ylabel('Qty (Pcs)')
plt.xlabel('Steps into the test data')
plt.show()

#Testing out a normal distribution for the max_features
import numpy as np
import scipy.stats as stats
import math
mu = ๐.๘
variance = ๐.๐๐๕
sigma = math.sqrt(variance)
x = np.linspace(mu - ๓*sigma, mu + ๓*sigma, ๑๐๐)
```

```
plt.plot(x, stats.norm.pdf(x, mu, sigma))
plt.show()

#Testing out a uniform distribution for the n_estimators
x = np.linspace(๐, ๒๐๐๐, ๑๐๐)
plt.plot(x, stats.uniform.pdf(x, ๕๐, ๙๕๐))
plt.show()

#RandomizedSearchCV with two distributions
from sklearn.model_selection import RandomizedSearchCV
#Specifying the distributions to draw from
distributions = {'max_features': stats.norm(๐.๘, math.sqrt(๐.๐๐๕)),'n_estimators':
stats.randint(๕๐, ๑๐๐๐)}
#Creating the search
my_rf = RandomizedSearchCV(RandomForestRegressor(), distributions, n_iter=๑๐,
scoring = 'r๒', n_jobs = -๑, random_state = ๑๒๓๔๕)
# Fitting the search
my_rf.fit(X_train, y_train)
rf_fcst = my_rf.predict(X_test)
print(my_rf.best_params_)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(rf_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.summary())
#RMSE
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(y_test, rf_fcst))
print("RMSE :\n", rmse)
#Feature importances
fi = pd.DataFrame({'feature': X_train.columns,'importance':
my_rf.best_estimator_.feature_importances_})
fi.sort_values('importance', ascending=False)
```

```
#KNN
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('train๑.csv') #filename
data.head()

#remove outliers
import numpy as np
Q๑ = np.percentile(data['Set-Qty'], ๒๕,
                   interpolation = 'midpoint')

Q๓ = np.percentile(data['Set-Qty'], ๗๕,
                   interpolation = 'midpoint')
IQR = Q๓ - Q๑

print("Old Shape: ", data.shape)

# Upper bound
upper = np.where(data['Set-Qty'] >= (Q๓+๑.๕*IQR))
# Lower bound
lower = np.where(data['Set-Qty'] <= (Q๑-๑.๕*IQR))

''' Removing the Outliers '''
data.drop(upper[๐], inplace = True)
data.drop(lower[๐], inplace = True)

print("New Shape: ", data.shape)

data = data.dropna()
X = data[['Year', 'Month', 'WeekDay', 'Setter', 'Metal', 'SetKind', 'StoneKind']]
y = data['Set-Qty']


# apply a min max scaler
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
```

```
data = pd.DataFrame(scaler.fit_transform(data), columns = data.columns)

# Create Train test split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=๐.๒,
random_state=๑๒๓๔๕, shuffle=True)
from sklearn.neighbors import KNeighborsRegressor
my_knn = KNeighborsRegressor()
my_knn.fit(X_train, y_train)
knn_fcst = my_knn.predict(X_test)
from sklearn.metrics import r๒_score
print(r๒_score(list(y_test), list(knn_fcst)))

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(knn_fcst, X_test๒).fit()
#est๒ = est.fit()
print("Adj. R๒ :\n",est.rsquared_adj)


#Creating a plot on the data of the test set
import matplotlib.pyplot as plt
plt.figure(figsize=(๒๐,๕))
plt.plot(list(y_test))
plt.plot(list(knn_fcst))
plt.legend(['actuals', 'forecast'])
plt.ylabel('Set-Qty')
plt.xlabel('Steps in test data')
plt.show()

#Adding a grid search cross-validation to the kNN model
from sklearn.model_selection import GridSearchCV
my_knn = GridSearchCV(KNeighborsRegressor(),
{'n_neighbors':[๒, ๔, ๖, ๘, ๑๐, ๑๒]},
scoring = 'r๒', n_jobs = -๑)
```

```
my_knn.fit(X_train, y_train)
knn_fcst = my_knn.predict(X_test)
print(r๒_score(list(y_test), list(my_knn.predict(X_test))))
print(my_knn.best_estimator_)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(knn_fcst, X_test๒).fit()
#est๒ = est.fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Adding a random search cross-validation to the kNN model
from sklearn.model_selection import RandomizedSearchCV
my_knn = RandomizedSearchCV(KNeighborsRegressor(),
{'n_neighbors':list(range(๑, ๒๐))},
scoring = 'r๒', n_iter=๑๐, n_jobs = -๑)
my_knn.fit(X_train, y_train)
knn_fcst = my_knn.predict(X_test)
print(r๒_score(list(y_test), list(my_knn.predict(X_test))))
print(my_knn.best_estimator_)
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(knn_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(knn_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.summary())
#RMSE
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(y_test, knn_fcst))
```

```
print("RMSE :\n", rmse)

#GradientBoosting
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv('train๔.csv')
data.head()

#reduce outliers
Q๑ = np.percentile(data['Set-Qty'], ๒๕, interpolation = 'midpoint')

Q๓ = np.percentile(data['Set-Qty'], ๗๕, interpolation = 'midpoint')
IQR = Q๓ - Q๑
print("Old Shape: ", data.shape)
# Upper bound
upper = np.where(data['Set-Qty'] >= (Q๓+๑.๕*IQR))
# Lower bound
lower = np.where(data['Set-Qty'] <= (Q๑-๑.๕*IQR))
''' Removing the Outliers '''
data.drop(upper[๐], inplace = True)
data.drop(lower[๐], inplace = True)
print("New Shape: ", data.shape)

data = data.dropna()
X = data[['Year', 'Month', 'WeekDay', 'Setter', 'Metal', 'SetKind', 'StoneKind']]
y = data['Set-Qty']

# apply a min max scaler
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data = pd.DataFrame(scaler.fit_transform(data), columns = data.columns)

# Create Train test split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=๐.๒,
```

```
random_state=๑๒๓๔๕, shuffle=True)
from xgboost import XGBRegressor
my_xgb = XGBRegressor()
my_xgb.fit(X_train, y_train)
xgb_fcst = my_xgb.predict(X_test)

#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(xgb_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Applying the default LightGBM model
from lightgbm import LGBMRegressor
my_lgbm = LGBMRegressor()
my_lgbm.fit(X_train, y_train)
lgbm_fcst = my_lgbm.predict(X_test)
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(lgbm_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Create a graph to compare the XGBoost and LightGBM forecast to the actuals
import matplotlib.pyplot as plt
plt.figure(figsize=(๒๐,๑๐))
plt.plot(list(y_test))
plt.plot(list(xgb_fcst))
plt.plot(list(lgbm_fcst))
plt.legend(['actual', 'xgb forecast', 'lgbm forecast'])
plt.ylabel('Traffic Volume')
plt.xlabel('Steps in test data')
plt.show()

#Applying Bayesian optimization to XGBoost
from skopt import BayesSearchCV
```

```
from skopt.space import Real, Categorical, Integer
import random
random.seed(๐)
xgb_opt = BayesSearchCV(
XGBRegressor(),
{
'learning_rate': (๑๐e-๖, ๑.๐, 'log-uniform'),
'max_depth': Integer(๐, ๕๐, 'uniform'),
'n_estimators' : (๑๐, ๑๐๐๐, 'log-uniform'),
},
n_iter=๑๐,
cv=๓
)
xgb_opt.fit(X_train, y_train)
xgb_tuned_fcst = xgb_opt.best_estimator_.predict(X_test)
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(xgb_tuned_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Applying Bayesian optimization to LightGBM
random.seed(๐)
lgbm_opt = BayesSearchCV(
LGBMRegressor(),
{
'learning_rate': (๑๐e-๖, ๑.๐, 'log-uniform'),
'max_depth': Integer(-๑, ๕๐, 'uniform'),
'n_estimators' : (๑๐, ๑๐๐๐, 'log-uniform'),
},
n_iter=๑๐,
cv=๓
)
lgbm_opt.fit(X_train, y_train)
lgbm_tuned_fcst = lgbm_opt.best_estimator_.predict(X_test)
#Adjust R๒
```

```
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(lgbm_tuned_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#Plotting the two tuned models
import matplotlib.pyplot as plt
plt.figure(figsize=(๒๐,๑๐))
plt.plot(list(y_test))
plt.plot(list(xgb_tuned_fcst))
plt.plot(list(lgbm_tuned_fcst))
plt.legend(['actual', 'xgb forecast', 'lgbm forecast'])
plt.ylabel('Traffic Volume')
plt.xlabel('Steps in test data')
plt.show()
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(xgb_tuned_fcst, X_test๒).fit()
#est๒ = est.fit()
print("Adj. R๒ :\n",est.rsquared_adj)

#result
#Adjust R๒
import statsmodels.api as sm
X_test๒ = sm.add_constant(X_test)
est = sm.OLS(lgbm_tuned_fcst, X_test๒).fit()
print("Adj. R๒ :\n",est.rsquared_adj)
#RMSE
from sklearn.metrics import mean_squared_error
from math import sqrt
rmse = sqrt(mean_squared_error(y_test, lgbm_tuned_fcst))
print("RMSE :\n", rmse)
```