การพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer โดยใช้การเรียนรู้ของเครื่อง
The scale predictions of multilayer PCB board by using machine learning

นัทธมน พลายอินทร์[1], ชนาพันธุ์ ชนาเนตร[2] ,วิกานดา ผาพันธ์[2*]
Nutthamon Plai-in[1], Chanaphun Chananet[2], Wikanda Phaphan[2*]

Received: 14 January 2020 ; Revised: 27 May 2020 ; Accepted: 10 July 2020

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer เพื่อกำหนดค่า Scale ของงาน เมื่อผ่านกระบวนการ Lamination Press แล้วทำให้แผ่นงานมีขนาดที่ได้ตรงตามความต้องการของลูกค้า โดยเก็บรวบรวมข้อมูลการกำหนดค่า Scale ของงานประเภท Multilayer จากบริษัทผลิตชิ้นส่วนอิเล็กทรอนิกส์แห่งหนึ่งย้อนหลัง ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2562 ข้อมูลที่ศึกษาประกอบไปด้วยตัวแปรตาม คือค่าเปอร์เซ็นต์ Scale ของงานชนิด Multilayer ตัวแปรอิสระ คือปัจจัยที่ส่งผลต่อการกำหนดค่าเปอร์เซ็นต์ Scale ของแผ่น PCB ชนิด Multilayer ได้แก่ Layer Count (จำนวนชั้นของงาน), Distance (ระยะของแนวงาน), Core Thickness (ความหนาของ Core), Thickness Cu Side (ความหนาของ Copper บนแผ่น Laminate Core นั้นๆ), Cu Foil (ความหนาของ Copper Foil), Tg (Glass Transition Temperature ของ Material), Cutdirection (แนวของงาน), Axis (แกนของงาน), Streak (แนวงาน Warp x Fill) และ Material Brand (ชนิดของแบรนด์ Material) ทำการศึกษาตัวแบบด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression : MLR) วิธีซัพพอร์ตเวกเตอร์ รีเกรสชัน (Support Vector Regression : SVR) และวิธีต้นไม้ตัดสินใจแบบรีเกรสชัน (Decision Tree Regression) ซึ่งเป็นตัวแบบของการเรียนรู้แบบมีผู้สอน (Supervised Learning) ในการเรียนรู้ของเครื่อง (Machine Learning) ประมวลผลโดยใช้โปรแกรม RStudio และเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์โดยพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE) ผลการวิจัยพบว่าตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเป็นตัวแบบที่ดีที่สุดและที่เหมาะสมกับข้อมูลการพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer โดยมีค่า MSE และค่า MAPE น้อยที่สุด ซึ่งช่วยเพิ่มความแม่นยำในการกำหนดค่า Scale และลดระยะเวลาในกระบวนการผลิตอันเนื่องมาจากความผิดพลาดในการกำหนดค่า Scale

คำสำคัญ: แผ่นพิมพ์วงจรอิเล็กทรอนิกส์, การถดถอยเชิงเส้นพหุคูณ, ซัพพอร์ตเวกเตอร์รีเกรสชัน, ต้นไม้ตัดสินใจแบบรี-เกรสชัน

Abstract

The purposes of this research were to predict and assign a scale value of multilayer PCB. After the PCB board underwent a process of lamination press, it made sheets be sized according to customers' needs. The collected data came from an electronics components manufacturing company in which there are the configuration data of the scale of the multilayer PCB board which went back from January 2018 to June 2019. The dependent variables were the percentage scale values of the multilayer PCB board and the independent variables were factors affecting the determination of the percentage scale values of the multilayer PCB board. Those factors were layer count, distance,

---

[1] นักศึกษา, ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ 1518 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพมหานคร 1080

[2] ผู้ช่วยศาสตราจารย์, ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ 1518 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพมหานคร 1080

[1] Student, Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.

[2] Assistant Professor, Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.

* Corresponding author: Tel: +66 931414615, E-mail: wikanda.p@sci.kmutnb.ac.th

core thickness, thickness Cu side (thickness of copper on laminate core board), thickness Cu Foil, Tg (glass transition temperature of material), cut direction, Axis, streak (Warp x Fill) and material brand. The models we used for analyzing were a multiple linear regression method (MLR), a support vector regression method (SVR), and a decision tree regression method. Those methods were the supervised learning models in machine learning and they were processed by using the RStudio program, mean square error (MSE) and mean absolute percent error (MAPE) for comparison of the efficiency of models for scale predictions. The result revealed that the MSE and the MAPE value of the support vector regression model are minimal, which means it is the most suitable model for the data of scale predictions of multilayer PCB board due to helping increase the accuracy in assigning scale values and helping save the time of the production process as a consequence of errors in assigning scale values.

Keyword: Printed circuit board, Multiple Linear Regression, Support Vector Regression, Decision Tree Regression

บทนำ

“แผ่นพิมพ์วงจรอิเล็กทรอนิกส์” หรือ PCB (PRINTED CIRCUIT BOARD) เป็นส่วนประกอบพื้นฐานที่สำคัญของวงจรอิเล็กทรอนิกส์ เป็นทางเดินสัญญาณไฟฟ้าให้แก่อุปกรณ์อิเล็กทรอนิกส์ต่างๆ ที่อยู่บนแผงวงจร ใช้สำหรับต่อวางอุปกรณ์อิเล็กทรอนิกส์เพื่อประกอบเป็นวงจรแทนการต่อวงจรด้วยสายไฟที่มีความซับซ้อนและยุ่งยาก ในปัจจุบันแผ่น PCB ถูกผลิตและนำไปใช้ในอุตสาหกรรมต่างๆ เช่น อุตสาหกรรมยานยนต์ อุปกรณ์โทรคมนาคมต่างๆ รวมถึง เครื่องมือทางการแพทย์ โทรศัพท์มือถือ และอุปกรณ์คอมพิวเตอร์

โดยทางบริษัทผลิตชิ้นส่วนอิเล็กทรอนิกส์ที่คณะผู้วิจัยศึกษา มีการผลิตแผ่นวงจรอยู่ 2 แบบ คือ การผลิตแบบ 2 หน้าเคลือบรู (Double Sided PCB) และการผลิตแบบหลายชั้น (Multilayer PCB) ในปัจจุบันบริษัทฯ มีความสามารถในการผลิตได้ตั้งแต่ 4 ถึง 24 ชั้น ซึ่งมีความซับซ้อนกว่าและใช้เทคโนโลยีสูงกว่าการผลิตแผ่นพิมพ์วงจรชนิด 2 หน้า กระบวนการผลิตแผ่น PCB ชนิด Multilayer จะต้องผ่านกระบวนการ Lamination Press ซึ่งเป็นการอัดงานให้ติดเป็นเนื้อเดียวกันโดยผ่านความร้อน ส่งผลให้ขนาดของงานมีการเปลี่ยนแปลง (Dimension Change) จึงต้องมีการกำหนด Scale ของงาน เพื่อทำการเผื่อระยะของแผ่นงาน ก่อนถึงกระบวนการ Lamination Press

ขั้นตอนการกำหนดค่า Scale ในปัจจุบันเป็นขั้นตอนที่ใช้ประสบการณ์ในการกำหนดค่า Scale และการเปรียบเทียบค่า Scale จากงานเก่าที่มีลักษณะโครงสร้างงานคล้ายกัน ซึ่งส่งผลให้ค่า Scale มีความแม่นยำค่อนข้างต่ำ และอาจเกิดการผิดพลาดจากการกำหนดค่า Scale ส่งผลให้ขนาดของแผ่นงานที่ได้ไม่ตรงตามที่ลูกค้ากำหนด จึงต้องมีการปรับค่า Scale ใหม่จนกว่าจะได้ขนาดของแผ่นงานที่เหมาะสมหรือเรียกว่าการลองผิดถูก (Trial and Error) ขั้นตอนดังกล่าวส่งผลให้เกิดความล่าช้าในการผลิตงาน ซึ่งถือเป็นการเสียโอกาส เสียเวลา และเสียค่าใช้จ่ายในการผลิตงานใหม่ให้กับลูกค้า ดังนั้นการกำหนดค่า Scale ให้เหมาะสมกับลักษณะโครงสร้างงานนั้นๆ ในระยะเวลาที่สั้น จะต้องอาศัยปัจจัยต่างๆ ที่เกี่ยวข้องกับการกำหนดค่า Scale เช่น จำนวนชั้นของงาน ความหนาของ Core และMaterial Brand เป็นต้น

จากความสำคัญและปัญหาที่กล่าวมาข้างต้น คณะผู้วิจัยจึงมีความสนใจศึกษา การถดถอยเชิงสถิติ (Statistical Regression) ของการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression : MLR) วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) และวิธีต้นไม้ตัดสินใจแบบรีเกรสชัน (Decision Tree Regression) หาตัวแบบที่เหมาะสมกับข้อมูลการกำหนดค่า Scale ของแผ่น PCB ชนิด Multilayer เพื่อให้การกำหนดค่า Scale มีความแม่นยำมากยิ่งขึ้น รวมถึงช่วยลดระยะเวลาและค่าใช้จ่ายในการผลิตงานใหม่ โดยใช้โปรแกรม RStudio เป็นเครื่องมือในการวิเคราะห์และหาตัวแบบพยากรณ์ โดยเกณฑ์ที่ใช้วัดความแม่นยำของตัวแบบได้แก่ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE)

วัตถุประสงค์การวิจัย

1. เพื่อเพิ่มความแม่นยำในการกำหนดค่า Scale ของงานชนิด Multilayer

2. เพื่อกำหนดค่า Scale ของแผ่น PCB ชนิด Multilayer อย่างเป็นระบบโดยใช้โปรแกรม RStudio ในการวิเคราะห์และพยากรณ์

3. เพื่อลดระยะเวลาในขั้นตอนการกำหนดค่า Scale

กระบวนการผลิตแผ่น PCB ที่เกี่ยวข้องกับการกำหนดค่า Scale

ขั้นตอนที่ 1 : Material Cutting

ขั้นตอนการตัดแผ่น Laminate ที่จะนำมาสร้างลายวงจร ต้องตัดให้มีขนาดตามที่แผนกออกแบบการผลิตได้กำหนดไว้

ขั้นตอนที่ 2 : Thin Core

การสร้างเส้นลายวงจรบนแผ่น Laminate มีขั้นตอนการปฏิบัติงานดังนี้

(1) Thin Core Chemical Clean

การเตรียมผิวทำความสะอาดผิวบอร์ดด้วยน้ำยาเคมี

(2) Thin Core Laminator

การนำงานที่ผ่านการทำความสะอาดแล้ว รีด Film คลุมทับทั้ง 2 ด้าน

(3) Thin Core Exposure

การถ่ายแสงเพื่อสร้างลายวงจร

(4) Develop

การนำงานที่ผ่านการถ่ายแสง ล้าง Film ส่วนที่ไม่ถูกการถ่ายแสงออก

(5) Etching

การกัดทองแดงส่วนที่ไม่ใช่ลายวงจรออกทั้งหมด

(6) Strip Film

การล้าง Film ส่วนที่คลุมเส้นลายวงจรออก หลังจากงานผ่านการ Strip Film จะได้งานที่มีลายวงจรชั้นใน

ขั้นตอนที่ 3 : Automatic Optical Inspection

การตรวจสอบคุณภาพลายวงจรเปรียบเทียบกับต้นแบบลูกค้า

ขั้นตอนที่ 4 : Lamination

(1) Optical Punch

การเจาะรูร้อยตาไก่ (Rivet)

(2) Oxide Treatment

การเคลือบผิวด้วยน้ำยาเคมี

(3) Ply Up และ Lay Up

การนำชิ้นงานมาเรียงประกอบให้เป็นตัวงาน

(4) Lamination Press

การนำตัวงานที่ผ่านการเรียงมา Press อัดให้งานติดเป็นเนื้อเดียวกันด้วยความร้อน

ขั้นตอนที่ 5 : X-Ray Drill

นำงานที่ผ่านการ Press แล้วมาเจาะหารูอ้างอิง

ตัวอย่างลักษณะการวางส่วนประกอบต่างๆ ของแผ่น PCB ชนิด Multilayer แสดงดัง Figure 1 และ 2

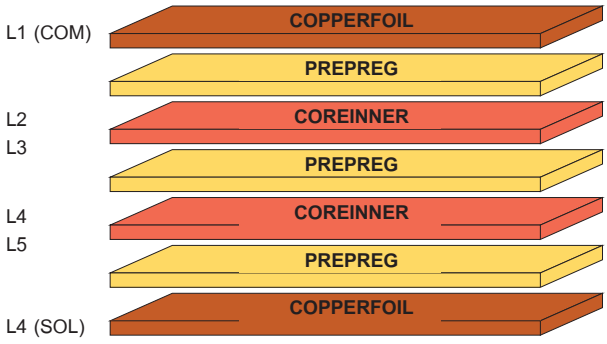

Figure 2 Multilayer 6 layer

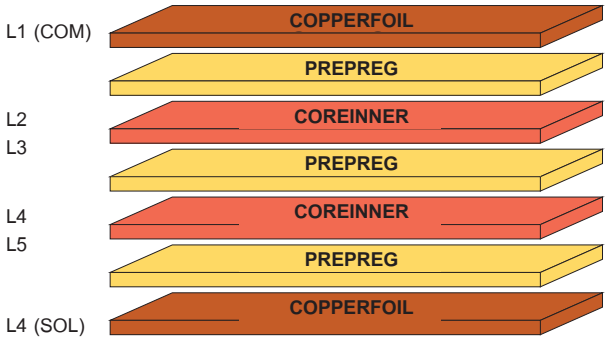

Figure 2 Multilayer 6 layer

Machine Learning

Machine Learning หรือ “การเรียนรู้ของเครื่อง” คือการที่ทำให้เครื่องจักรสามารถคำนวณทำตามชุดคำสั่งได้หรือทำให้มีความสามารถในการเรียนรู้ด้วยตนเอง โดยการเรียนรู้และพยายามเข้าใจรูปแบบความสัมพันธ์ข้อมูลงานใดงานหนึ่งเป็นตัวอย่างและเมื่อมีข้อมูลใหม่เข้ามาสามารถทำนายหรือตัดสินใจจากตัวอย่างด้วยตนเอง ซึ่งเป็น AI (Artificial Intelligent : ปัญญาประดิษฐ์) อีกอัลกอริทึมหนึ่งที่กำลังได้รับความนิยม ง่ายต่อการเขียนโปรแกรมภาษา การประยุกต์ใช้งานผู้ใช้จะเขียนโปรแกรมภาษาให้คอมพิวเตอร์เรียนรู้ด้วยตนเอง โดยป้อนข้อมูลให้กับโปรแกรมภาษาที่เขียนไว้ให้เรียนรู้เองจากข้อมูลที่ป้อนไปเรื่อยๆ จนถึงจุดที่โปรแกรมภาษาที่เขียนไว้เริ่มจำแนกได้แม่นยำแล้วจึงเอามาใช้งาน โดยการเรียนรู้ของเครื่องสามารถแบ่งประเภทได้เป็น 2 ประเภทหลักใหญ่ๆ[1-4] ได้แก่

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)

สามารถนำไปประยุกต์ใช้แก้ปัญหาได้ 2 รูปแบบ คือ การถดถอยเชิงสถิติ (Statistical Regression) และการแบ่งประเภท (Classification)

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

โดยส่วนใหญ่จะนำไปประยุกต์ใช้แก้ปัญหาในเรื่องของการแบ่งกลุ่ม (Cluster)

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) เป็นการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์และการเปลี่ยนแปลงของตัวแปรตาม (Dependent Variable) หรือตัวแปรตอบสนอง (Response Variable) ที่มีอิทธิพลเนื่องมาจากตัวแปรอิสระ (Independent Variable) หรือตัวแปรทำนาย (Predictor Variable) ตั้งแต่ 2 ตัวแปรขึ้นไป โดยอาศัยความสัมพันธ์เชิงเส้นตรงระหว่างกลุ่มของตัวแปรอิสระที่มีต่อตัวแปรตามมาใช้ในการประมาณค่าหรือทำนายค่าของตัวแปรตามได้ [5]

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$$

เมื่อ

Y คือตัวแปรตาม

β_0 คือค่าคงที่ (Constant) ของสมการถดถอย

β_i คือค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระ เมื่อ $i = 1, 2, ..., k$

x_i คือตัวแปรอิสระที่ i

k คือจำนวนตัวแปรอิสระในสมการถดถอย

ε คือความคลาดเคลื่อนสุ่ม (Random Error)

ซัพพอร์ตเวกเตอร์รีเกรสชัน

ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) ได้แนวคิดและพัฒนามาจาก ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) เพื่อนำไปประยุกต์ใช้สำหรับงานด้านการประมาณค่าหรือการวิเคราะห์การถดถอย (Regression) ซึ่งใน SVM จะใช้สมการระนาบเกิน (Hyperplane) ทำการแบ่งข้อมูลออกจากกัน แต่ SVR นั้นจะใช้ Hyperplane ในการพยากรณ์ค่า[6-8] โดยมีลักษณะสมการเป็นดังนี้

$$f(x) = (w \bullet x) + b \quad (1)$$

เมื่อ

w คือเวกเตอร์ถ่วงน้ำหนัก (Weight)

b คือค่าความคลาดเคลื่อน (Bias)

x คือเวกเตอร์ข้อมูลนำเข้า (Input data) แสดงในรูปของเวกเตอร์คอลัมน์ $(x_1, ..., x_n)^T$ ขนาดเท่ากับ n ; $x_i \in R$; $i = 1, ..., n$

R คือเซตของจำนวนจริง

หลักการในการหา Hyperplane สำหรับใช้แทนกลุ่มของข้อมูลที่เหมาะสมที่สุด จะหาตำแหน่งของข้อมูลที่เป็นซัพพอร์ตเวกเตอร์ (Support Vector) ซึ่งจะเป็นข้อมูลในตำแหน่งที่อยู่ห่างจาก Hyperplane มากที่สุด และตำแหน่งดังกล่าวจะอยู่บนเส้นแบ่งระยะของเขต (Boundary Line) ที่กำหนดขึ้น

ซึ่งในทางทฤษฎีจะพยายามให้ข้อมูลทั้งหมดอยู่ภายใน Boundary Line และสร้าง Hyperplane ที่ใช้แทนกลุ่มข้อมูลขึ้น เพื่อใช้ประมาณค่าให้ใกล้เคียงกับตัวอย่างข้อมูลเอ้าพุต โดย Hyperplane ดังกล่าวจะอยู่ในตำแหน่งที่รักษาระยะห่างที่มากที่สุด ระหว่างตำแหน่งข้อมูลที่เป็น Support Vector

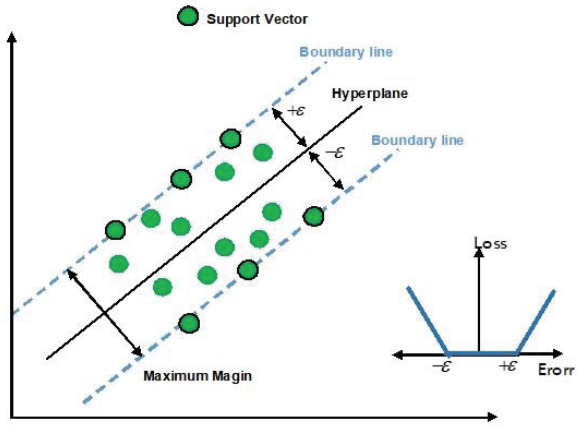

Figure 3 Show how to find the optimal hyperplane line to represent the data group[9]

สำหรับ SVR กรณีปัญหาแบบไม่เชิงเส้น (Non-Linear) ไม่สามารถใช้วิธีแบบเชิงเส้น (Linear) ดังนั้นจึงต้องใช้ฟังก์ชันเคอร์เนล (Kernel Function) ในการส่งผ่าน (Mapping) ข้อมูลที่ไม่เป็นเชิงเส้นเพื่อให้สามารถใช้วิธีการแบบเชิงเส้นได้ ซึ่งประเภทของ Kernel Function นั้นมีหลายประเภท เช่น Linear Kernel Function, Sigmoid Kernel Function และGaussian Radial Basis (RBF) เป็นต้น

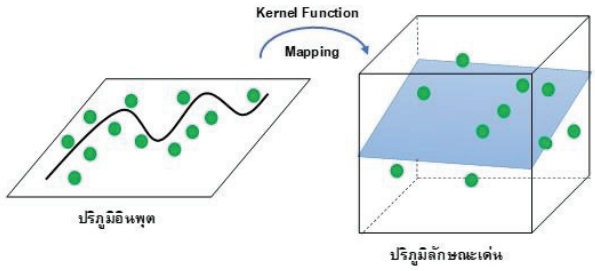

Figure 4 Mapping Kernel Function in Support Vector Regression

ฟังก์ชันเคอร์เนลที่นิยมใช้ในการสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน คือ Gaussian Radial Basis Function[7]

$$k(x,x_i) = \exp(-\gamma \| x - x_i \|^2), \gamma > 0 \qquad (2)$$

ดังนั้นสมการ Hyperplane สามารถเขียนใหม่ในรูปแบบสมการระนาบเกินไม่เป็นเชิงเส้นโดยใช้ฟังก์ชันเคอร์เนลได้ดังนี้

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \qquad (3)$$

ต้นไม้ตัดสินใจแบบรีเกรสชัน

Decision Tree หรือเทคนิคต้นไม้ตัดสินใจ คือแบบจำลองทางคณิตศาสตร์ เพื่อหาทางเลือกที่ดีที่สุด โดยนำข้อมูลมาสร้างการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งสามารถแบ่งออกเป็น 2 ประเภทหลักๆ คือ Decision Tree Classification (ต้นไม้สำหรับการจัดกลุ่ม) และ Decision Tree Regression (ต้นไม้สำหรับการทำนายค่า) โดยโครงสร้างของต้นไม้ตัดสินใจจะประกอบไปด้วย

- โหนดราก คือ ตำแหน่งเริ่มต้นของต้นไม้ตัดสินใจ

- โหนดกิ่ง คือ โหนดลูก มีจำนวนโหนดขึ้นอยู่กับ Condition ของโหนดแม่

- โหนดใบ คือ ผลลัพธ์ที่ต้องการในการทำนายของตัวแบบพยากรณ์

ในกรณี Decision Tree Regression เป็นเทคนิคการทำให้ผลลัพธ์เป็นค่าจำนวนจริงหรือเป็นผลลัพธ์ที่ต่อเนื่องกัน (Continuous Output) ของสิ่งที่ต้องการพยากรณ์ โดยหลักการสร้างจะเริ่มสร้างจากโหนดรากและแบ่งข้อมูลออกเป็น 2 เส้นทาง โดยใช้กฎ If-else ดัง Figure 5 ในการแบ่งต้นไม้ และทำการแตกโหนดไปเรื่อยๆ จนถึงโหนดใบ จะทำการพยากรณ์ค่าผลลัพธ์ที่ต้องการ ด้วยการหาค่าเฉลี่ยของโหนดผลลัพธ์และค่าเฉลี่ยที่ได้จะเป็นผลลัพธ์ของการพยากรณ์[10-13]

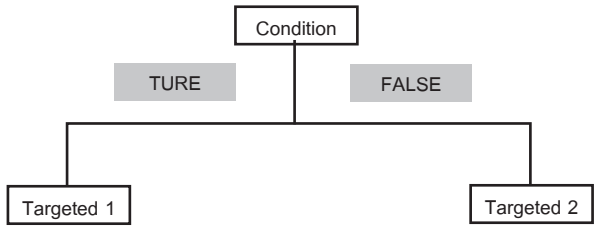

Figure 5 Show if-else statements used to divide the tree[11]

ตัวอย่าง จาก Figure 6 เมื่อทำการแบ่งข้อมูลจะเห็นได้ว่า ถ้าข้อมูล $k_1 \le t_1$ และ $X_2 \le t_2$ เป็นจริงทั้ง 2 เงื่อนไข ผลลัพธ์ของการพยากรณ์คือ R_1 โดยที่ R_1 คือค่าเฉลี่ยของโหนดผลลัพธ์

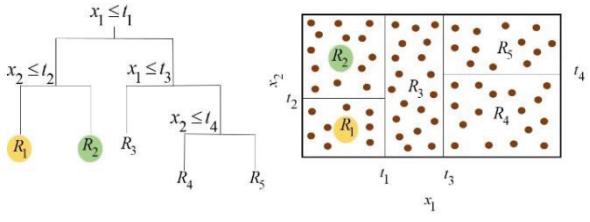

Figure 6 Previewing the data division of the Decision Tree Regression

วิธีดำเนินการวิจัย

ในการดำเนินการครั้งนี้ ผู้วิจัยได้ทำการศึกษาในเรื่องการสร้างตัวแบบเพื่อใช้พยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer ซึ่งมีกระบวนการดำเนินการวิจัยประกอบด้วย 6 ขั้นตอนหลัก คือ

ขั้นตอนที่ 1 : การออกแบบโครงงานวิจัย

งานวิจัยนี้มีรูปแบบการดำเนินงานโดยเก็บข้อมูลจากฐานข้อมูลในบริษัท ในลักษณะของการเก็บข้อมูลย้อนหลัง (Retrospective) เพื่อศึกษาหารูปแบบความสัมพันธ์ระหว่างปัจจัยต่างๆ ที่ส่งผลต่อการกำหนดค่า Scale และหาค่าตัวแบบเพื่อใช้พยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer

ขั้นตอนที่ 2 : ตัวแปรที่ใช้ในโครงงานวิจัย

1. ตัวแปรต้น (Independent Variable)

- ตัวแปรเชิงปริมาณ (Quantitative Variable)

(1) Layer Count คือจำนวนชั้นของงาน มีตั้งแต่ 4-24 ชั้น มีหน่วยเป็น layer

(2) Distance คือระยะของแนวงาน มีหน่วยเป็น inches

(3) Core thickness คือความหนาของ Core มีหน่วยเป็น mill

(4) Thickness Cu Side คือความหนาของ Copper บนแผ่น Laminate Core นั้นๆ มีหน่วยเป็น Ounce

(5) Cu Foil คือความหนาของ Copper Foil ที่นำมาใช้ Press มีหน่วยเป็น Ounce

(6) Tg คือGlass Transition Temperature ของ Material มีหน่วยเป็น Degree Celsius (°C)

- ตัวแปรเชิงคุณภาพ (Qualitative Variable)

(1) Cut direction คือแนวของงานมี 2 แบบ คือ Cut A และ Cut B ไม่มีหน่วย

(2) Axis คือ แกนของงานมี 2 แกน คือ แกน X และแกน Y ไม่มีหน่วย

(3) Streak คือแนวของงาน Warp x Fill ไม่มีหน่วย

(4) Material Brand คือชนิดของแบรนด์ Material ไม่มีหน่วย

2. ตัวแปรตาม (Dependent Variable) ได้แก่ ค่า %Scale X และ %Scale Y

ขั้นตอนที่ 3 : การรวบรวมข้อมูล (Data Collection)

การวิจัยครั้งนี้เป็นการวิเคราะห์ข้อมูลเพื่อหาแบบจำลองทางสถิติ โดยเก็บจากแหล่งข้อมูลทุติยภูมิ (Secondary Data) ซึ่งเป็นข้อมูลที่รวบรวมจากฐานข้อมูลของบริษัทผลิตชิ้นส่วนอิเล็กทรอนิกส์ที่คณะผู้วิจัยศึกษา โดยนำข้อมูลรายวัน ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2562 มาวิเคราะห์หาตัวแบบในการพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer

ขั้นตอนที่ 4 : การเตรียมข้อมูล (Data Preparation)

กระบวนการในการเตรียมข้อมูลมีความสำคัญกับงานด้านการทำนาย สาเหตุที่ต้องมีการเตรียมข้อมูล คือ ข้อมูลมีค่าสูญหาย (Missing Value) หมายถึง ข้อมูลที่จัดเก็บบางส่วนอาจเป็นค่าว่าง ข้อมูลมีค่าผิดปกติ (Outlier) หมายถึง ข้อมูลมีค่ามากกว่าหรือน้อยกว่าค่าที่ควรจะเป็น หรืออาจเกิดจากความผิดพลาดในการจดบันทึกข้อมูลและข้อมูลที่ได้ไม่มีความสม่ำเสมอ (Inconsistent) หมายถึง ข้อมูลอาจอยู่ในรูปแบบต่างกันหรืออาจมีหน่วยวัดที่ต่างกัน สามารถแก้ไขได้โดยการเตรียมข้อมูล ซึ่งแบ่งออกเป็น 3 ส่วน ดังนี้

1. การคัดเลือกข้อมูล (Data Selection) ทำการคัดเลือกข้อมูลที่เกี่ยวข้องกับการวิเคราะห์ และเป็นประโยชน์ต่อการพยากรณ์ สิ่งที่ผู้วิจัยสนใจศึกษาเป็นเรื่องของการกำหนดค่า Scale ซึ่งเลือกใช้ตัวแปรที่ระบุไว้ในขั้นตอนที่ 2

2. การกลั่นกรองข้อมูล (Data Cleaning) หลังจากคัดเลือกข้อมูลแล้ว พบว่าข้อมูลยังไม่สมบูรณ์ เช่น ค่าสูญหาย (Missing Value) และมีการเก็บข้อมูลซ้ำในฐานข้อมูล แก้ไขโดยการตัดแถวนั้นทิ้งไป โดยใช้โปรแกรม RStudio ในการแก้ไข

3. การแปลงรูปแบบข้อมูล (Data Transformation)

3.1 การจัดรูปแบบการเก็บข้อมูล

เมื่อคัดกรองข้อมูลที่ไม่ถูกต้องออกจากชุดข้อมูลแล้ว ได้ทำการจัดรูปแบบการเก็บข้อมูลใหม่ โดยการแบ่งข้อมูลของแต่ละ Part Number ตามจำนวน Core Inner ของงานนั้นๆ และทำการแยกข้อมูลของแต่ละ Core Inner ด้วยแกนของงาน (แกน X และแกน Y) ทำให้ใน 1 Core Inner ของ Part Number จะแบ่งข้อมูลออกเป็น 2 บรรทัด ซึ่งในการจัดรูปแบบดังกล่าวจะทำให้มีข้อมูลในการวิเคราะห์มากขึ้น และสามารถพยากรณ์ค่า Scale ในทีเดียวได้ทั้งแกน X และแกน Y

3.2 การ Encoding ตัวแปรเชิงคุณภาพ

เนื่องจากงานวิจัยนี้มีตัวแปรเชิงคุณภาพ ยังไม่สามารถนำมาวิเคราะห์ เพื่อสร้างตัวแบบสำหรับการพยากรณ์ได้ จึงได้ทำการ Encode กับตัวแปรเชิงคุณภาพ โดยใช้วิธีการ Encode แบบ One-Hot Encoding คือ การแปลงข้อมูลให้เป็น Binary (0 กับ 1) โดยใช้คำสั่ง dummy_cols() จาก Packages “fastDummies” ในโปรแกรม RStudio

ขั้นตอนที่ 5 : การสร้างตัวแบบ (Modeling)

ผู้วิจัยได้ดำเนินการสร้างตัวแบบเพื่อพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer ด้วยตัวแบบการพยากรณ์ 3 ตัวแบบ ได้แก่ ตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) และตัวแบบต้นไม้ตัดสินใจแบบรีเกรสชัน (Decision Tree Regression)

ขั้นตอนที่ 6 : การเปรียบเทียบประสิทธิภาพของตัวแบบ (Evaluation)

ในงานวิจัยนี้ได้ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE) เปรียบเทียบค่าเพื่อเลือกตัวแบบที่ดีที่สุดในการพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer

ผลการวิจัย

ตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression : MLR)

โปรแกรม RStudio มีฟังก์ชันพื้นฐานสำหรับคำนวณสถิติ คือ ANOVA และ Linear Regression Model ซึ่งใช้กับกรณีตัวแปรตาม (Dependent Variable) ที่เป็นข้อมูลต่อเนื่อง โดย Linear Regression ใช้เมื่อตัวแปรต้น (Independent Variable) เป็นข้อมูลตัวเลขต่อเนื่อง สามารถใช้คำสั่ง lm( ) คำนวณตัวแบบ กำหนดให้ตัวแปรตามคือ ค่า Scale และตัวแปรต้นคือตัวแปรที่มีความสัมพันธ์กับตัวแปรตาม โดยใช้คำสั่งในโปรแกรม RStudio ช่วยในการสร้างตัวแบบ ดัง Figure 7 ได้สมการถดถอยดังนี้

$$
\begin{aligned}
Scale\% = {} & 1.568x10^{-3} + 2.269x10^{-3}LayerCount \\
& -2.054x10^{-5}Distance + 7.586x10^{-5}Corethk \\
& -3.468x10^{-3}ThkcusideA + 9.164x10^{-3}ThkcusideB \\
& -3.548x10^{-4}Cufoil + 2.263x10^{-5}Tg \\
& -2.248x10^{-4}CutDirectionA \\
& -1.800x10^{-3}AxisX + 8.590x10^{-3}Streakfill \\
& -7.158x10^{-3}BrandTLM140MT \\
& -2.023x10^{-3}BrandTLM150MT \\
& +5.071x10^{-3}BrandTLM170T \\
& +3.897x10^{-3}BrandTLM170TF2 \\
& -2.258x10^{-3}BrandTLM150MTF2 \\
& -3.227x10^{-3}BrandPANASONICR1566W \\
& -1.682x10^{-3}BrandTLM140 \\
& -1.876x10^{-3}BrandSHENGYIS1000 \\
& +1.503x10^{-3}BrandTLM150HF \\
& +1.952x10^{-3}BrandSHENGYIS1000.2 \\
& +4.418x10^{-4}BrandSHENGYIS1000AUTOLAD1 \\
& +9.221x10^{-3}BrandMATSUSHITAR1566W \\
& +3.959x10^{-3}BrandNANYANP175RBH \\
& +1.813x10^{-3}BrandAUTOLAD3 \\
& -9.522x10^{-3}BrandTLM140MTREV1 \\
& +2.742x10^{-3}BrandISOLA370HR
\end{aligned}
$$

โดยที่

Scale% คือ ค่าเปอร์เซ็นต์ Scale ของงานชนิด Multilayer

LayerCount คือ จำนวนชั้นของงาน

Distance คือ ระยะของแนวงาน

Corethk คือ ความหนาของ Core

ThkcusideA คือ ความหนาของ Cu บนแผ่น Laminate Core นั้นๆ หน้า A

ThkcusideB คือ ความหนาของ Cu บนแผ่น Laminate Core นั้นๆ หน้า B

Cufoil คือ ความหนาของ Copper Foil ที่นำมาใช้ Press

Tg คือ Glass Transition Temperature ของ Material

CutdirectionA คือ แนวของงาน Cut A

AxisX คือ แกน X ของงาน

Streakfill คือ แนวของงานด้าน fill

BrandTLM140MT คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 1

BrandTLM150MT คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 2

BrandTLM170T คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 3

BrandTLM170TF2 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 4

BrandTLM150TF2 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 5

BrandPANASONICR1566W คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 6

BrandTLM140 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 7

BrandSHENGYIS1000 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 8

BrandTLM150HF คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 9

BrandSHENGYIS1000.2 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 10

BrandSHENGYIS1000AUTOLAD1 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 11

BrandMATSUSHITAR1566W คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 12

BrandNANYANP175RBH คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 13

BrandAUTOLAD3 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 14

BrandTLM140MTREV1 คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 15

BrandISOLA370HR คือ ชื่อชนิดของแบรนด์ที่ใช้ในการผลิตแบบที่ 16

```
####################fit multiple linear regression###################
regressor1 = lm(formula = Scale ~ .,data = dataset )
```

Figure 7 Commands in RStudio to create Multiple Linear Regression

ตรวจสอบข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

1. การตรวจสอบการแจกแจงปกติของค่าความคลาดเคลื่อน ได้ผลลัพธ์ Figure 8 พบว่า กราฟที่ได้ไม่มีแนวโน้มเป็นเส้นตรง สรุปได้ว่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

2. การตรวจสอบค่าความแปรปรวนและความเป็นอิสระของค่าความคลาดเคลื่อน ได้ผลลัพธ์ดัง Figure 9 พบว่า ค่าความคลาดเคลื่อนมีการกระจายอยู่รอบค่าศูนย์แบบสุ่มละขนานไปกับแกน X แสดงว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่ และค่าความคลาดเคลื่อนเป็นอิสระกัน

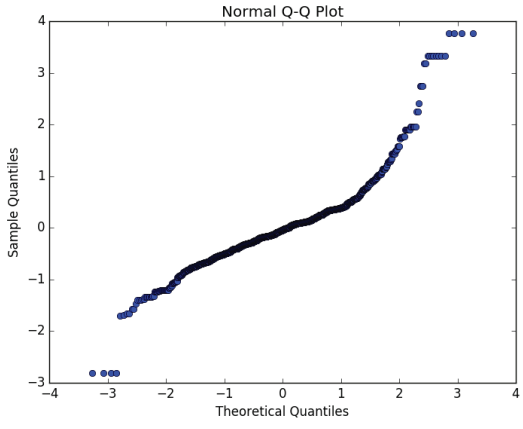

Figure 8 Normal Quantile-Quantile Plot of Regression Standardized Residual of the scale forecast of multilayer PCB board

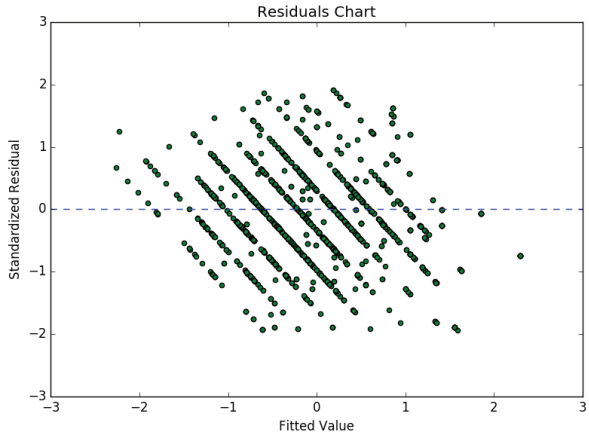

Figure 9 The relationship between the standardized residual and the predictions of scale values of multilayer PCB board

ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR)

การหาตัวแบบพยากรณ์ Support Vector Regression ในโปรแกรม RStudio สามารถทำได้โดยใช้คำสั่ง svm( ) จาก Packages “e1071” โดยตั้งค่าให้ type = ‘eps-regression’ กำหนดค่าพารามิเตอร์เป็นค่าเริ่มต้น (Default) ของโปรแกรม และกำหนดให้ตัวแปรตามคือ ค่า Scale และตัวแปรต้นคือตัวแปรที่เหลือ ดัง Figure 10

```
####################Fitting the SVR Model to the dataset####################
library(e1071)
regressor2 = svm(formula = Scale ~.,
                 data = dataset,
                 type = 'eps-regression')
```

Figure 10 Commands in RStudio to create Support Vector Regression

ตัวแบบต้นไม้ตัดสินใจแบบรีเกรสชัน (Decision Tree Regression)

การหาตัวแบบพยากรณ์ Decision Tree Regression ในโปรแกรม RStudio สามารถทำได้โดยใช้คำสั่ง rpart( ) จาก Packages “rpart” โดยตั้งค่าให้ control = rpart.control (minsplit=1) กำหนดให้พารามิเตอร์เป็นค่าเริ่มต้น (Default) ของโปรแกรม และกำหนดให้ตัวแปรตามคือ ค่า Scale และตัวแปรต้นคือตัวแปรที่เหลือ ดัง Figure 11

```
##########Fitting the decision tree to the dataset##########

library(rpart)
regressor3 = rpart(formula = Scale ~ .,
                   data = dataset,
                   control = rpart.control(minsplit = 1))
```

Figure 11 Commands in RStudio to create Decision Tree Regression

เปรียบเทียบค่า MSE และ MAPE

จากการพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer ทั้ง 3 ตัวแบบได้ผลลัพธ์ดัง Table 1

Table 1 The MSE and the MAPE values of all 3 predictive models

	MSE	MAPE (%)
Multiple Linear Regression	0.00006	23.89107
Support Vector Regression	0.00004	17.52377
Decision Tree Regression	0.00015	45.84676

วิจารณ์และสรุปผล

การดำเนินงานครั้งนี้ ผู้วิจัยทำการสร้างตัวแบบเพื่อพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer โดยใช้ข้อมูลการกำหนดค่า Scale ซึ่งเป็นข้อมูลที่เกี่ยวกับโครงสร้างของงานแต่ละ Part Number ของบริษัทผลิตชิ้นส่วนอิเล็กทรอนิกส์แห่งหนึ่ง ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2562 รวม 17 เดือน โดยมีปัจจัยที่นำเข้าไปใช้ในการวิเคราะห์การพยากรณ์ค่า Scale มีจำนวน 10 ตัวแปร ได้แก่ Layer Count, Distance, Core Thickness, Thickness Cu Side, Cu Foil, Tg, CutDirection, Axis, Streak และ Material Brand เมื่อจัดการเตรียมข้อมูลเสร็จสมบูรณ์ได้นำไปวิเคราะห์สร้างตัวแบบโดยใช้ตัวแบบจากการเรียนรู้แบบมีผู้สอน (Supervised Learning) ประเภทการถดถอยเชิงสถิติ (Statistics Regression) ในการเรียนรู้ของเครื่อง (Machine Learning) ได้ทำการเลือกมา 3 ตัวแบบคือ ตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) และตัวแบบต้นไม้ตัดสินใจแบบรีเกรสชัน (Decision Tree Regression) โดยใช้โปรแกรม RStudio เป็นเครื่องมือในการประมวลผลวิเคราะห์และหาตัวแบบการพยากรณ์ แล้วทำการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) เพื่อตรวจสอบว่าตัวแบบที่สร้างออกมานั้นมีความถูกต้องหรือผิดพลาดมากน้อยเพียงใด มีผลสรุปดัง Figure 12

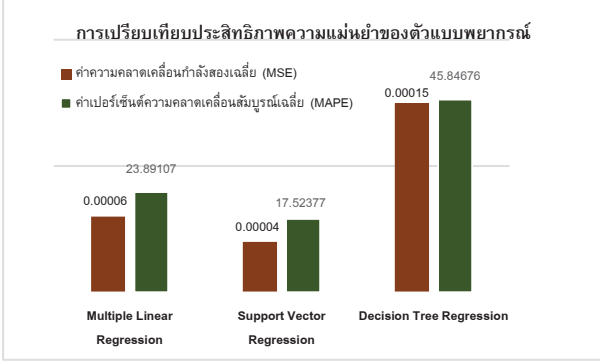

Figure 12 Comparison of efficiency and accuracy of 3 predictive models

ผลการวิจัยพบว่าจากตัวแบบการพยากรณ์ทั้ง 3 ตัวแบบนั้น ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) มีค่าความคลาดเคลื่อนน้อยที่สุด โดยมีค่า MSE = 0.00004 และค่า MAPE = 17.52377% รองลงมาคือตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ มีค่า MSE = 0.00006 และค่า MAPE = 23.89107% แต่เนื่องจากตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณไม่ผ่านสมมติฐานของความคลาดเคลื่อน คือค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ ดังนั้นตัวแบบการวิเคราะห์การถดถอยเชิงเส้นพหุคูณจึงไม่เหมาะที่จะใช้ พยาการณ์ จึงสรุปว่าตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเป็นตัวแบบที่เหมาะสมกับข้อมูลการพยากรณ์ค่า Scale ของแผ่น PCB ชนิด Multilayer มากที่สุด

จากนั้นผู้วิจัยได้จัดทำโปรแกรมภาษา ที่สามารถเรียนรู้ได้ด้วยตนเอง พร้อมสื่อการเรียนรู้ในรูปแบบเอกสารคู่มือการใช้โปรแกรม RStudio เพื่อพยากรณ์ค่า Scale ของงานชนิด Multilayer สำหรับพนักงานของบริษัทผลิตชิ้นส่วนอิเล็กทรอนิกส์ที่คณะผู้วิจัยศึกษา เพื่อให้สามารถนำตัวแบบไปประยุกต์ใช้งานด้วยโปรแกรม RStudio ได้

เอกสารอ้างอิง

1. Piasak A. Machine Learning. [อินเทอร์เน็ต]. 2560 [เข้าถึงเมื่อ 10 สิงหาคม 2562]. เข้าถึงได้จาก: https://blog.derlivery.com/machine-learning-b390451983d0
2. G-Able. Data Science ก้าวที่ล้ำหน้ากว่า Bid Data (จบ). [อินเทอร์เน็ต]. [เข้าถึงเมื่อ 10 สิงหาคม 2562]. เข้าถึงได้จาก: https://www.g-able.com/digital-review/digital-transformation/big-data-analytics/data-science-%E0%B8%81%E0%B9%89%E0%B8%B2%E0%B8%A7%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%A5%E0%B9%89%E0%B8%B3%E0%B8%AB%E0%B8%99%E0%B9%89%E0%B8%B2%E0%B8%81%E0%B8%-A7%E0%B9%88%E0%B8%B2-big-data-2/
3. Kean J. สอนคอมพูดแบบไทยๆ ด้วย Machine Learning Model [Part 2]. [อินเทอร์เน็ต]. 2562 [เข้าถึงเมื่อ 10 กรกฎาคม 2562]. เข้าถึงได้จาก: https://medium.com/@newnoi/%E0%B8%AA%E0%B8%AD%E0%B8%99%E0%B8%84%E0%B8%AD%E0%B8%A1%E0%B8%9E%E0%B8%B9%E0%B8%94%E0%B9%81%E0%B8%9A%E0%B8%9A%E0%B9%84%E0%B8%97%E0%B8%A2%E0%B9%86-%E0%B8%94%E0%B9%89%E0%B8%A7%E0%B8%A2-machine-learning-model-part2-2a1609af1bd7
4. Voungchuy S. Machine Learning: 01-Data Preprocessing (python coding[basic]). [อินเทอร์เน็ต]. 2561 [เข้าถึงเมื่อ 12 กันยายน 2562]. เข้าถึงได้จาก: https://medium.com/@sirawit0676/machine-learning-01-data-preprocessing-python-coding-basic-687aee03c478
5. ธิดาเดียว มยุรีสวรรค์. การวิเคราะห์การถดถอย Regression Analysis. พิมพ์ครั้งที่ 1. ขอนแก่น: บริษัทเพ็ญพรินติ้ง จำกัด ; 2559.
6. จารุมน หนูคง. การศึกษาเปรียบเทียบเทคนิคการพยากรณ์ข้อมูลอนุกรมเวลาราคายางพาราด้วยวิธีโครงข่ายประสาทเทียม สมการถดถอยแบบโพลิโนเมียล และซัพพอร์ตเวกเตอร์รีเกรสชัน. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ; 2552.
7. ธีร์ธวัช แก้ววิจิตร. การเพิ่มประสิทธิภาพซัพพอร์ตเวกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา. นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุร-นารี ; 2559.
8. รณชัย ชื่นธวัช. การพยากรณ์ความต้องการใช้งานหน่วยจำหน่ายไฟฟ้าด้วยซัพพอร์ตเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน. วิทยาศาสตร์และเทคโนโลยี. 2560 ; 19(1) : 217-218.
9. รัชภูมิ ใจกล้า, พันธุ์ทิพย์ นนทรีย์, อรรถชัย จินตะเวช. การทำนายผลผลิตข้าวโดยวิธีซัพพอร์ตเวกเตอร์แมชีน. ใน: รายงานการประชุมวิชาการ ศวพก. ปี 2550. เชียงใหม่: ศูนย์วิจัยเพื่อเพิ่มผลผลิตทางเกษตร ; 2550. หน้า 10-15.
10. วิชญ์พงษ์ ดรุณธรรม. รู้จัก Decision Tree, Random Forest, และ XGBoost!!! - PART 1. [อินเทอร์เน็ต]. 2561 [เข้าถึงเมื่อ 5 กันยายน 2562]. เข้าถึงได้จาก: https://medium.com/@witchapongdaroontham/%E0%-B8%A3%E0%B8%B9%E0%B9%89%E0%B8%88%E0%B8%B1%E0%B8%81-decision-tree-random-forrest-%E0%B9%81%E0%B8%A5%E0%B8%B0-xgboost-part-1-cb49c4ac1315
11. พัชราวรรณ ชินไธสง. การออกแบบและพัฒนาเทคนิคไฮบริดสำหรับการเติมค่าข้อมูลที่สูญหาย. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์. นครราชสีมา: บัณฑิตวิทยาลัยมหาวิทยาลัยเทคโนโลยีสุรนารี ; 2555.
12. อโณทัย ศิลเทพาเวทย์. แบบจำลองเพื่อพัฒนาคุณภาพของผลิตภัณฑ์เอชจีเอในโรงงานอุตสาหกรรมฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ.วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย ; 2554.
13. ฤทธิไกร ทวีเจริญ. ตัวแบบการพยากรณ์การลาออกของพนักงานฝ่ายผลิตในธุรกิจเครื่องฉีดพลาสติกของบริษัทแห่งหนึ่ง.วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชานโยบายและการบริหารเทคโนโลยีสารสนเทศ. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์ ; 2559.