

การเปรียบเทียบประสิทธิภาพเทคนิคการเรียนรู้ของเครื่องสำหรับการบำรุงรักษาเชิงคาดการณ์สำหรับเครื่องยนต์อากาศยาน

The Comparison of Machine Learning Algorithms for Predictive Maintenance of Aircraft Engine

วिरากานต์ กิตติบวรกุล¹, ศรายุทธ นนท์ศิริ², พิชิตชัย คำอินทร์³

ห้องปฏิบัติการวิจัยปัญญาประดิษฐ์และอินเทอร์เน็ตของสรรพสิ่ง

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย-ญี่ปุ่น,

Wirakarn Kittibawornkul¹, Sarayut Nonsiri², Pichitchai Kamin³

AIoT Research Laboratory, Faculty of Information Technology, Thai-Nichi Institute of Technology

ki.wirakarn_st@tni.ac.th¹, sarayut.n@tni.ac.th², pichitchai@tni.ac.th³

Received: April 28, 2022; Revised: May 18, 2022; Accepted: May 30, 2022

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลการบำรุงรักษาเชิงคาดการณ์ ของเครื่องยนต์อากาศยานด้วยเทคนิคการเรียนรู้ของเครื่อง และเปรียบเทียบประสิทธิภาพในการคาดการณ์ความเสียหายของแต่ละเทคนิคเพื่อหาเทคนิคที่มีความเหมาะสม โดยใช้ชุดข้อมูลจากโมเดลใหม่ของ Commercial Modular Aero-Propulsion System Simulation (N-CMAPSS) ในการฝึกฝนและทดสอบโมเดลในการคาดการณ์ความเสียหาย ชุดข้อมูลดังกล่าวเป็นชุดข้อมูลเกี่ยวกับความเสียหายที่เกิดขึ้นกับเครื่องยนต์อากาศยานแบบ Turbofan จากศูนย์ความเป็นเลิศด้านการทำนาย (The Prognostics Center of Excellence) ของศูนย์วิจัยนาซาเอมส์ (NASA Ames Research Center) เทคนิคการเรียนรู้ของเครื่องทั้งหมด 5 เทคนิคได้แก่ ต้นไม้ตัดสินใจ ป่าแบบสุ่ม ขั้นตอนวิธีเพื่อนบ้านใกล้สุด โครงข่ายประสาทเทียม และ Gradient Boosting with XGBoost ถูกนำมาใช้สำหรับสร้างโมเดลเพื่อคาดการณ์ความเสียหาย ในส่วนของการประเมินโมเดลใช้ค่าสัมประสิทธิ์การกำหนด (R^2) และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE) จากการฝึกฝนและทดสอบโมเดลทั้งหมด 5 เทคนิคพบว่า เทคนิคโครงข่ายประสาทเทียมให้ผลลัพธ์ที่ดีที่สุดด้วยค่าสัมประสิทธิ์การกำหนดเท่ากับ 0.7578 และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ยเท่ากับ 0.0016

คำสำคัญ: การเรียนรู้ของเครื่อง การคาดการณ์ความเสียหาย การบำรุงรักษาเชิงคาดการณ์ เครื่องยนต์อากาศยาน ชุดข้อมูล N-CMAPSS

ABSTRACT

This paper aimed to create the predictive maintenance model of an aircraft engine by implementing the machine learning techniques and comparing the prediction performance of each technique to determine the most suitable technique. Datasets for this study were collected from a new model of Commercial Modular Aero-Propulsion System Simulation (N-CMAPSS) used for training and testing for the predictive maintenance model. N-CMAPSS was

a series of data on the damage to turbofan engines from The Prognostics Center of Excellence (PCoE), NASA Ames Research Center. There were 5 machine learning techniques including Decision Tree, Random Forest, k-Nearest Neighbor (kNN), Artificial Neural Network (ANN), and Gradient Boosting with XGBoost. They were used to implement the predictive maintenance model. In addition, the designation coefficient of determination (R^2) and Root mean square error (RMSE) were used to evaluate the model. The analysis of the training and testing of five techniques concluded that Artificial Neural Network was the suitable technique with 0.7578 of R^2 and 0.0016 of RMSE.

KEYWORDS: Machine Learning, Failure Prediction, Predictive Maintenance, Aircraft Engine, N-CMAPSS Datasets

บทนำ

เครื่องยนต์เป็นระบบกลไกที่มีความสำคัญต่อการขับเคลื่อนของอากาศยานด้วยโครงสร้างที่มีความซับซ้อนและจำเป็นที่จะต้องทำงานอยู่ในสถานะที่ไม่เอื้ออำนวย ส่งผลให้มีโอกาสที่ประสิทธิภาพของเครื่องยนต์จะลดลงซึ่งหมายถึงค่าใช้จ่ายที่เพิ่มสูงขึ้นในหลายด้าน ด้วยอัตราการสิ้นเปลืองเชื้อเพลิงที่มากขึ้นรวมถึงค่าบำรุงรักษาที่อาจตามมาในกรณีที่เกิดการเสื่อมสภาพและอาจนำไปสู่ความเสียหายที่เป็นอันตรายอย่างยิ่ง ดังนั้นความสมบูรณ์ของชิ้นส่วนเครื่องยนต์และการทำงานที่มีความเชื่อถือได้สูงจึงเป็นสิ่งสำคัญอย่างยิ่งของความปลอดภัยขณะทำการบิน (Junjie, Jinqian, & Feng, 1707) (Efstratios & Pantelis, 2015) ด้วยเหตุนี้ความสำคัญของการบำรุงรักษาจึงเพิ่มขึ้นและส่งผลกระทบต่อแนวโน้มความสนใจในการพัฒนารวมถึงการปรับใช้กลยุทธ์การบำรุงรักษาที่มีประสิทธิภาพเพิ่มสูงขึ้นเช่นกัน การบำรุงรักษาที่มีประสิทธิภาพจะสามารถปรับปรุงความน่าเชื่อถือของระบบป้องกันความล้มเหลวของระบบ และลดต้นทุนการบำรุงรักษาได้ (Ana, et al., 2018) อีกทั้งในช่วงที่ผ่านมาการบำรุงรักษาเชิงคาดการณ์ (Predictive Maintenance) ได้รับความนิยมสูงขึ้น การปฏิวัติอุตสาหกรรม 4.0 ก็ให้การสนับสนุนการบำรุงรักษาประเภทนี้จากความต้องการในภาคปฏิบัติที่เพิ่มสูงขึ้นเช่นกัน เนื่องจากมีความได้เปรียบในเรื่องของความน่าเชื่อถือ ความพร้อม

ในการใช้งาน และความปลอดภัยที่มากกว่าการบำรุงรักษาแบบเดิม (Khanh & Kamal, 2019) ซึ่งหมายถึงการบำรุงรักษาเชิงรับ (Corrective Maintenance) ที่จะทำให้การซ่อมบำรุงเมื่อเกิดความเสียหายขึ้น เป็นวิธีการที่มีใช้มาอย่างยาวนานและเป็นที่นิยมมากจากความง่ายในการดำเนินการ แต่ก็จะแฝงไปด้วยค่าใช้จ่ายที่มากมาย ไม่ว่าจะเป็นค่าใช้จ่ายที่ต้องเสียไปขณะที่กระบวนการผลิตหยุดชะงักหรือค่าดำเนินงานต่าง ๆ รวมถึงการซ่อมบำรุงเชิงป้องกัน (Preventive Maintenance) ที่เป็นการซ่อมบำรุงตามแผนที่ได้วางไว้ แนวทางนี้จะสามารถป้องกันความเสียหายที่อาจเกิดขึ้นได้แต่ก็จะเป็นการใช้ทรัพยากรสิ้นเปลืองโดยใช่เหตุและเสียค่าใช้จ่ายที่ไม่จำเป็นอย่างมหาศาลเช่นกัน (Gian, Andrea, Simone, & Sean, 2015) จากการศึกษาพบว่ามีสามแนวทางหลักที่ใช้สำหรับคาดการณ์ความเสียหายที่อาจเกิดขึ้นได้แก่แนวทางโมเดลทางกายภาพ (Physical Model-Based) ซึ่งจะใช้โมเดลทางคณิตศาสตร์และวิธีการทางสถิติเป็นหลัก แนวทางความรู้ (Knowledge-Based) เป็นแนวทางที่จะช่วยลดความซับซ้อนของโมเดลทางกายภาพจึงนิยมใช้แบบผสมผสาน (Hybrid) และแนวทางการใช้ข้อมูล (Data-Driven) มีการนำโมเดลที่หลากหลายมาปรับใช้ในการพัฒนาการบำรุงรักษาเชิงคาดการณ์โดยพื้นฐานทางสถิติ (Statistics-Based) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) และโมเดลที่อ้างอิง

จากเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ถือเป็นที่ยอมรับอย่างมากในปัจจุบัน (Tiago, et al., 2020) การเรียนรู้ด้วยเครื่อง ถือเป็นเครื่องมือที่มีสมรรถภาพสูงในการพัฒนาอัลกอริทึมเพื่อประยุกต์ใช้สำหรับการคาดการณ์มากมาย โดยประสิทธิภาพของการนำไปประยุกต์ใช้นั้นขึ้นอยู่กับทางเลือกเทคนิคการเรียนรู้ของเครื่องให้มีความเหมาะสม (Thyago, et al., 2019) จากเหตุผลข้างต้นการเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องจะสามารถชี้ให้เห็นว่าชุดข้อมูล (Dataset) ที่มีอยู่นั้นเหมาะสมที่จะใช้เทคนิคไหนมาคาดการณ์ความเสียหายได้ แต่ด้วยข้อจำกัดในทางปฏิบัติ การที่จะรวบรวมข้อมูลความเสียหายประเภทที่มีการใช้งานจนกระทั่งเกิดความเสียหาย (Run-To-Failure) ของเครื่องยนต์อากาศยานนั้นเป็นไปได้ยาก เนื่องจากต้องใช้ระยะเวลา มีค่าใช้จ่ายสูงมาก (Abhinav, Kai, Don, & Neil, 2008) และเหตุผลสำคัญด้านความปลอดภัยที่ส่งผลกระทบต่ออย่างร้ายแรงหากเกิดเหตุการณ์เช่นนั้นขึ้นจึงมีการซ่อมบำรุงก่อนเสมอ ดังนั้นงานวิจัยนี้จะนำชุดข้อมูลที่สร้างขึ้นจากโมเดลใหม่ ของ Commercial Modular Aero-Propulsion System Simulation (N-CMAPSS) (Manuel, Chetan, Kai, & Olga, 2021) มาใช้สำหรับฝึกฝนโมเดลที่มีการใช้เทคนิคการเรียนรู้ของเครื่องที่แตกต่างกันทั้งหมด 5 เทคนิคและทำการเปรียบเทียบผลลัพธ์ที่ได้เพื่อหาโมเดลที่มีความเหมาะสมสำหรับใช้ในการคาดการณ์ความเสียหายของชิ้นส่วนอากาศยานต่อไป

วัตถุประสงค์

1. เพื่อสร้างโมเดลการบำรุงรักษาเชิงคาดการณ์ของเครื่องยนต์อากาศยานด้วยเทคนิคการเรียนรู้ของเครื่อง
2. เพื่อเปรียบเทียบประสิทธิภาพด้านความแม่นยำในการคาดการณ์ของแต่ละเทคนิค และหาเทคนิคที่มีความเหมาะสมที่สุด

ขอบเขตงานวิจัย

1. ฝึกฝนและทดสอบโมเดลที่ใช้ในการคาดการณ์ความเสียหายของเครื่องยนต์อากาศยานด้วยชุดข้อมูลที่ได้จาก Commercial Modular Aero-Propulsion System Simulation (N-CMAPSS)
2. เปรียบเทียบประสิทธิภาพด้านความแม่นยำของโมเดลการบำรุงรักษาเชิงคาดการณ์ของเครื่องยนต์อากาศยาน โดยใช้เทคนิคดังต่อไปนี้
 - (1) ต้นไม้ตัดสินใจ (Decision Tree)
 - (2) ป่าแบบสุ่ม (Random Forest)
 - (3) ขั้นตอนวิธีเพื่อนบ้านใกล้สุด (k-Nearest Neighbor: kNN)
 - (4) โครงข่ายประสาทเทียม (Artificial Neural Network: ANN)
 - (5) Gradient Boosting with XGBoost

ประโยชน์ที่ได้รับ

1. โมเดลสำหรับการคาดการณ์ความเสียหายของเครื่องยนต์อากาศยานในแต่ละเทคนิค
2. การตัดสินใจเลือกใช้เทคนิคการเรียนรู้ของเครื่องให้มีความเหมาะสมกับชุดข้อมูล
3. นำความรู้ด้านการเรียนรู้ของเครื่องมาประยุกต์ใช้ในการบำรุงรักษาเชิงคาดการณ์อื่น
4. นำความรู้ที่ได้ไปพัฒนาการบำรุงรักษาเชิงคาดการณ์ให้มีประสิทธิภาพมากยิ่งขึ้น

งานวิจัยที่เกี่ยวข้อง

(เสนหา, 2559) ได้ทำการศึกษาวิจัยเกี่ยวกับการประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network: CNN) ในการทำนายอายุการใช้งานคงเหลือของเครื่องจักร ซึ่งมีการนำชุดข้อมูลที่ได้จากเครื่องจำลองระบบการขับเคลื่อนทางการบินเชิงพาณิชย์ (Commercial Modular Aero-Propulsion System Simulation: CMAPSS) มาใช้ร่วมกับชุดข้อมูลอื่นรวม 14 ชุด โดยผลลัพธ์ที่ได้

แสดงให้เห็นว่าวิธี Enhanced VGG-16 ให้ค่าความคลาดเคลื่อนในการทำนายที่น้อยที่สุด

(Vimala, Tom, Vikram, B , & M , 2017) ทำการศึกษาเปรียบเทียบโมเดลจากเทคนิคการเรียนรู้ของเครื่องทั้งหมด 10 เทคนิคเพื่อหาเทคนิคการเรียนรู้ของเครื่องที่สามารถใช้ทำนายอายุการใช้งานคงเหลือ (Remaining Useful Lifetime: RUL) ของเครื่องยนต์อากาศยานแบบ Turbofan ได้อย่างแม่นยำมากที่สุด โดยผลลัพธ์ที่ได้แสดงให้เห็นว่าเทคนิคป่าแบบสุ่มเป็นเทคนิคที่ให้ความแม่นยำมากที่สุดจากการประเมินด้วยค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE)

(Rounak & Manikandan, 2021) มีจุดมุ่งหมายที่จะนำเทคนิค Long Short-Term Memory (LSTM) แบบที่มีโครงข่ายประสาทเทียมคอนโวลูชัน (Convolution Neural Network: CNN) และไม่มีโครงข่ายประสาทเทียมคอนโวลูชันมาใช้ในการคาดการณ์อายุการใช้งานคงเหลือ (RUL) ด้วยชุดข้อมูล CMAPSS ซึ่งผลลัพธ์ที่ได้แสดงให้เห็นว่า LSTM แบบที่ไม่มีโครงข่ายประสาทเทียมคอนโวลูชันนั้นดีกว่า

(Nathaniel , Christopher, & Katherine , 2021) มุ่งเน้นการใช้โมเดลเกี่ยวกับข้อมูล (Data-based models) เพื่อทำการหาอายุการใช้งาน

คงเหลือ (RUL) ของเครื่องยนต์แบบ Turbofan ด้วยชุดข้อมูลใหม่ของ CMAPSS หรือ N-CMAPSS ซึ่งโครงข่ายประสาทเทียมคอนโวลูชันถูกนำมาศึกษาในการหาอายุการใช้งานคงดังกล่าว

วิธีดำเนินการวิจัย

ชุดข้อมูล

ชุดข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นชุดข้อมูลเกี่ยวกับความเสียหายที่เกิดขึ้นกับเครื่องยนต์อากาศยานจากระบบฐานข้อมูล Standard Datasets ของ NASA (Chao, 2022) ถูกสร้างขึ้นจากโปรแกรม Commercial Modular Aero-Propulsion System Simulation หรือ C-MAPSS ซึ่งเป็นโปรแกรมสำหรับจำลองข้อมูลเครื่องยนต์อากาศยานเชิงพาณิชย์ขนาดใหญ่ที่มีลักษณะเครื่องยนต์แบบ Turbofan ในส่วนของการวิจัยฉบับนี้จะใช้ชุดข้อมูลที่ได้จากโมเดลที่ปรับปรุงใหม่ในชื่อ N-CMAPSS มาทำการเปรียบเทียบการคาดการณ์ความเสียหายของชิ้นส่วนเครื่องจักรด้วยโมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน โดยโมเดลของระบบจะเป็นไปตามสมการที่ 1 รายละเอียดเกี่ยวกับตัวแปรต่างๆ สามารถศึกษาได้จากงานวิจัย (Chao, 2022)

$$[\hat{x}_s^{(t)}, \hat{x}_v^{(t)} = S(\omega^{(t)}, \theta^{(t)})] \quad (1)$$

เมื่อ $\hat{x}_s^{(t)}$ คือคุณสมบัติทางกายภาพจากการวัด
 $\hat{x}_v^{(t)}$ คือคุณสมบัติจากเซนเซอร์เสมือนจริง
 $\omega^{(t)}$ คือข้อบ่งชี้การทำงานของเครื่องยนต์
 $\theta^{(t)}$ คือพารามิเตอร์แสดงสถานะการเสื่อมสภาพ

การเตรียมชุดข้อมูล

ในขั้นตอนของการเตรียมชุดข้อมูลจะเป็นการเตรียมข้อมูลเพื่อเข้าสู่กระบวนการวิเคราะห์ หากข้อมูลนั้นไม่มีความสมบูรณ์หรือไม่ได้มีการ

เตรียมชุดข้อมูลให้มีความเหมาะสมจะส่งผลให้การสร้างโมเดลเกิดปัญหาหรือเกิดข้อผิดพลาดในการวิเคราะห์ได้อาทิเช่น ค่าผิดปกติ (Outliers) ในชุดข้อมูลจะส่งผลให้ ความแม่นยำของโมเดลลดลงหรือเกิดข้อผิดพลาดมากขึ้น โดยกระบวนการสำคัญคือการทำความสะอาดข้อมูล (Clean Data)

1. การจัดการข้อมูลสูญหาย (Missing Data) เมื่อข้อมูลที่ได้มาอาจไม่มีความสมบูรณ์ด้วยเหตุปัจจัยหลายอย่าง จึงจำเป็นที่จะต้องหาวิธีจัดการกับข้อมูลที่สูญหายไปเหล่านี้ เพื่อให้เกิดความผิดพลาดในการวิเคราะห์น้อยที่สุด โดยหลัก

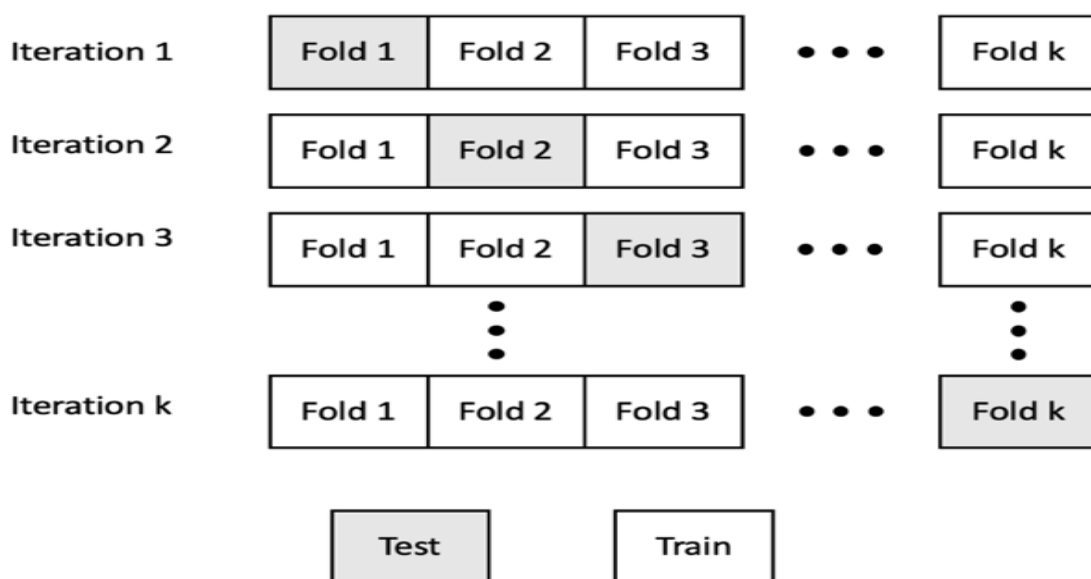
แล้วจะแบ่งออกเป็นสองวิธีได้แก่ การลบทิ้ง และการแทนที่

2. การจัดการชุดข้อมูลเป็นคำกล่าวถึงภาพรวมของการทำความสะอาดข้อมูล นอกเหนือจากการจัดการกับชุดข้อมูลสูญหาย ซึ่งจะมีขั้นตอนวิธีหลากหลายสำหรับการจัดการให้ชุดข้อมูลมีความสมบูรณ์ พร้อมสำหรับการนำไปวิเคราะห์หรือสร้างโมเดล ไม่ว่าจะเป็นการลบคอลัมน์ที่ไม่ต้องการ การเปลี่ยนชื่อคอลัมน์ให้เหมาะสม การจัดการชนิดของข้อมูล (Data Type) การตรวจสอบข้อมูลซ้ำ และขั้นตอนอื่น ๆ ตามความเหมาะสม

3. การจัดการค่าผิดปกติ (Outliers) เป็นการจัดการกับค่าที่มีความแตกต่างจากค่าอื่นในชุด

ข้อมูลอย่างผิดปกติที่อาจมีสาเหตุจากหลายปัจจัย ซึ่งเป็นค่าที่ไม่เหมาะสมที่จะนำไปใช้เนื่องจากจะก่อให้เกิดความผิดพลาดกับโมเดลหรือการวิเคราะห์ได้

4. การตรวจสอบไขว้ (Cross-Validation) เป็นการแบ่งชุดข้อมูลที่มีออกเป็นข้อมูลฝึกฝน (Training Data) และข้อมูลทดสอบ (Testing Data) ซึ่งนิยมแบ่งในอัตราส่วน 80:20 หรือ 70:30 ตามลำดับ จากนั้นจะจัดกลุ่มชุดข้อมูลออกเป็น k ชุดหรือที่เรียกว่า k-Fold Cross-Validation เพื่อเป็นการลดความเอนเอียงต่อข้อมูล เนื่องด้วยวิธีการนี้ข้อมูลทั้งหมดจะมีโอกาสเป็นทั้งข้อมูลฝึกฝนและข้อมูลทดสอบ ผลลัพธ์ที่ได้จะเป็นค่าเฉลี่ยค่าจากตรวจสอบทั้งหมด k รอบดังรูปที่ 1



ภาพประกอบที่ 1 k-Fold Cross-Validation

ในส่วนของงานวิจัยนี้ไม่พบข้อมูลค่าว่าง (Null) แต่ได้มีการจัดการกับข้อมูลเพื่อแบ่งเป็นตัวแปรต้นหรือตัวแปรอิสระ (x) ซึ่งเป็นค่าที่ได้จากการวัดด้วยเซนเซอร์ต่าง ๆ และตัวแปรตาม (y) ซึ่งเป็นค่าพารามิเตอร์แสดงสถานะการเสื่อมสภาพ โดยมีการแบ่งข้อมูลฝึกฝนจำนวน 552,088 แถวและข้อมูลทดสอบจำนวน 136,263 แถว

การสร้างโมเดล

ในขั้นตอนนี้เทคนิคตามที่ได้กล่าวไว้จะถูกนำมาใช้ในการสร้างโมเดลด้วยภาษา Python บนโปรแกรม Visual Studio Code และจะมีการเรียกใช้ Library ดังต่อไปนี้

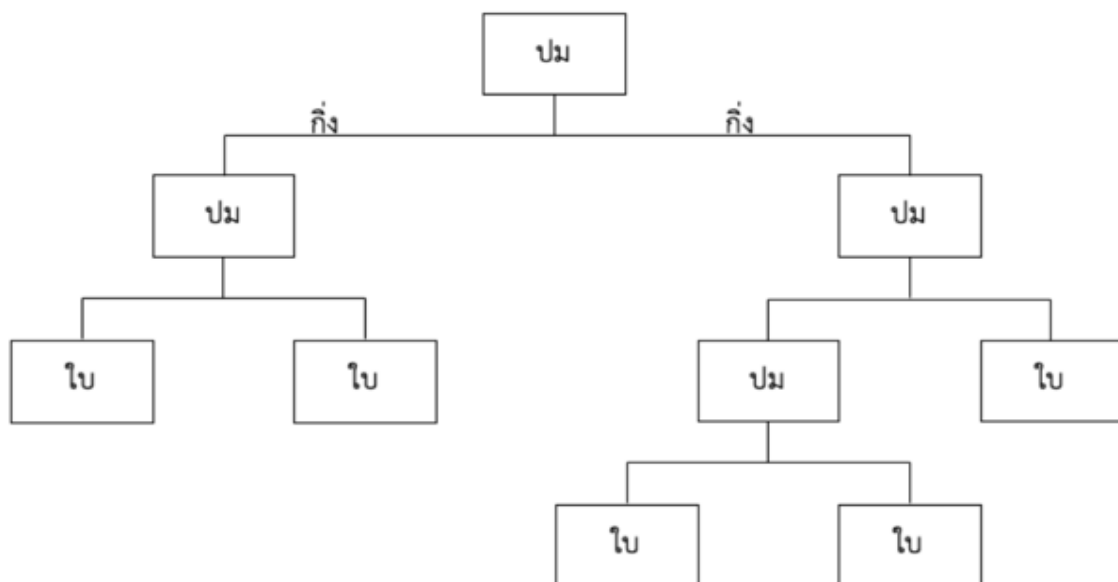
1. Pandas สำหรับการจัดการข้อมูล
2. h5py สำหรับอ่านไฟล์ .h5
3. Numpy สำหรับการคำนวณทางคณิตศาสตร์และสถิติ

4. Math สำหรับการใช้ฟังก์ชันทางคณิตศาสตร์
5. Scikit-Learn สำหรับสร้างโมเดลการเรียนรู้ของเครื่อง
6. Keras สำหรับสร้างโมเดลโครงข่ายประสาทเทียม

เพื่อทำการคาดการณ์ความเสียหายของชิ้นส่วนเครื่องจักรด้วยชุดข้อมูลที่ได้เตรียมไว้ในสองขั้นตอนที่ผ่านมาและเพื่อการเปรียบเทียบประสิทธิภาพของโมเดลนั้นรายละเอียดในแต่ละเทคนิคจะถูกกล่าวถึงต่อจากนี้

1. ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคที่เป็นที่รู้จักและมีความแพร่หลายเนื่องจากมีรูปแบบที่เข้าใจง่าย มีลักษณะการเรียนรู้แบบมี

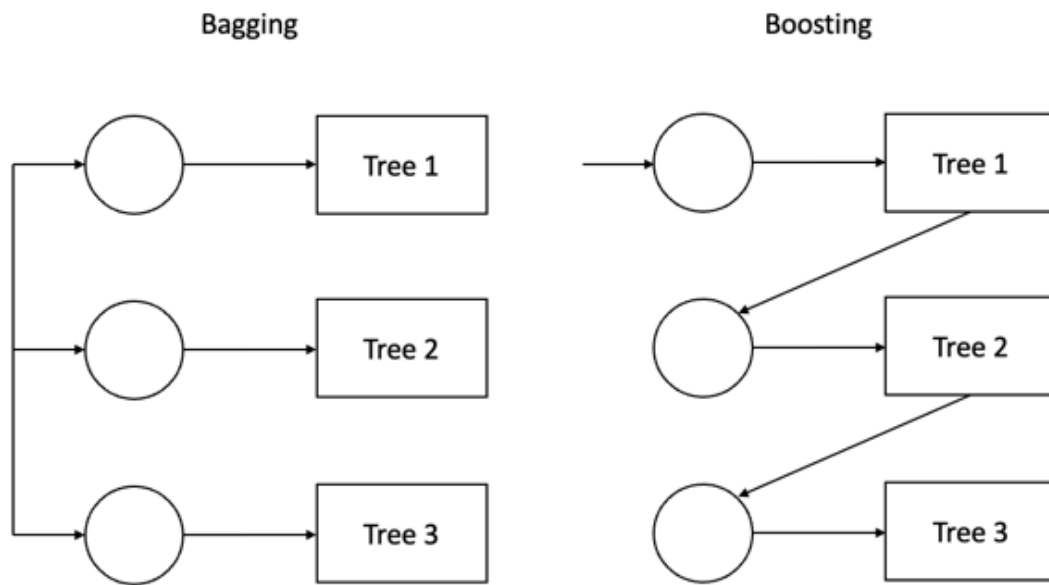
ผู้สอนสำหรับแก้ปัญหาได้ทั้งการจำแนกประเภท (Classification) ที่ผลลัพธ์จะเป็นกลุ่มหรือหมวดหมู่ (Categorical) และการถดถอย (Regression) ที่ตัวแปรจะเป็นแบบต่อเนื่อง (Continuous) แต่ส่วนใหญ่จะนิยมใช้สำหรับการจำแนกประเภท ต้นไม้ตัดสินใจมีองค์ประกอบหลักสามอย่างด้วยกันได้แก่ ปม (Node) กิ่ง (Branch) และใบ (Leaf) ในแต่ละปมจะแสดงถึงลักษณะประจำ (Attributes) ในหมวดหมู่และจะมีปมราก (Root Node) เป็นปมแรก การตัดสินใจหรือผลลัพธ์ที่ได้จะถูกแทนด้วยใบ ซึ่งจะมีกิ่งสำหรับแสดงค่าของแต่ละคุณลักษณะดังรูปที่ 2 (J. R., 1986)



ภาพประกอบที่ 2 ต้นไม้ตัดสินใจ

2. ป่าแบบสุ่ม (Random Forest) ถึงแม้ว่าต้นไม้ตัดสินใจจะสามารถเข้าใจได้ง่ายแต่ก็พบว่ามีแนวโน้มที่จะไม่มีความเสถียร (Unstable) ควบคุมขนาดได้ยาก และอาจเจอปัญหาที่โมเดลขณะฝึกสอนมีประสิทธิภาพดีกว่าความเป็นจริง (Overfitting) จึงได้มีการพัฒนาอัลกอริทึมป่าแบบสุ่มมาเพื่อแก้ปัญหานี้ (J. R., 1986) ป่าแบบสุ่มจัดเป็นเทคนิคการเรียนรู้ของเครื่องที่พัฒนามาจากต้นไม้ตัดสินใจถือเป็นแนวทางการเรียนรู้แบบ Ensemble คือการ

รวมหลายโมเดลเข้าด้วยกันสำหรับแก้ปัญหาที่มีความซับซ้อน สามารถใช้ได้ทั้งการจำแนกประเภทและการถดถอย แบ่งออกเป็นสองประเภทได้แก่ Bagging ซึ่งเป็นการเรียนรู้แบบแยกจากกันโดยอิสระแล้วจึงนำมาประมวลผลในตอนสุดท้ายและ Boosting ซึ่งเป็นการเรียนรู้แบบที่จะทำการเรียนรู้และปรับปรุงแบบจำลองไปที่ละขั้นตอนดังรูปที่ 3 (Leo, 2001)



ภาพประกอบที่ 3 ปาแบบสุม

3. ขั้นตอนวิธีเพื่อนบ้านใกล้สุด (k-Nearest Neighbor: kNN) เป็นเทคนิคที่ได้รับความนิยมเป็นอย่างมากในกลุ่มของการจำแนกประเภท เนื่องจากเป็นวิธีที่สามารถเข้าใจได้ง่ายไม่ซับซ้อนและมีประสิทธิภาพสูง ไม่เพียงแต่การจำแนกประเภทเท่านั้นที่เทคนิคนี้สามารถใช้ได้ แต่ยังรวมถึงกลุ่มของการถดถอยเช่นกัน โดยปกติ

ในงานของการจำแนกประเภทเทคนิคนี้จะทำการจัดกลุ่มของข้อมูลใหม่ให้อยู่ในกลุ่มเดียวกับข้อมูลที่ใกล้เคียง ค่า k จะเป็นค่าที่กำหนดจำนวนชุดข้อมูลที่ใกล้เคียงที่ต้องการใช้ในการจำแนกประเภท แต่ในส่วนของการถดถอยนั้นจะเป็นไปตามสมการ Euclidean distance (Korstanje, The kNN Model, 2021)

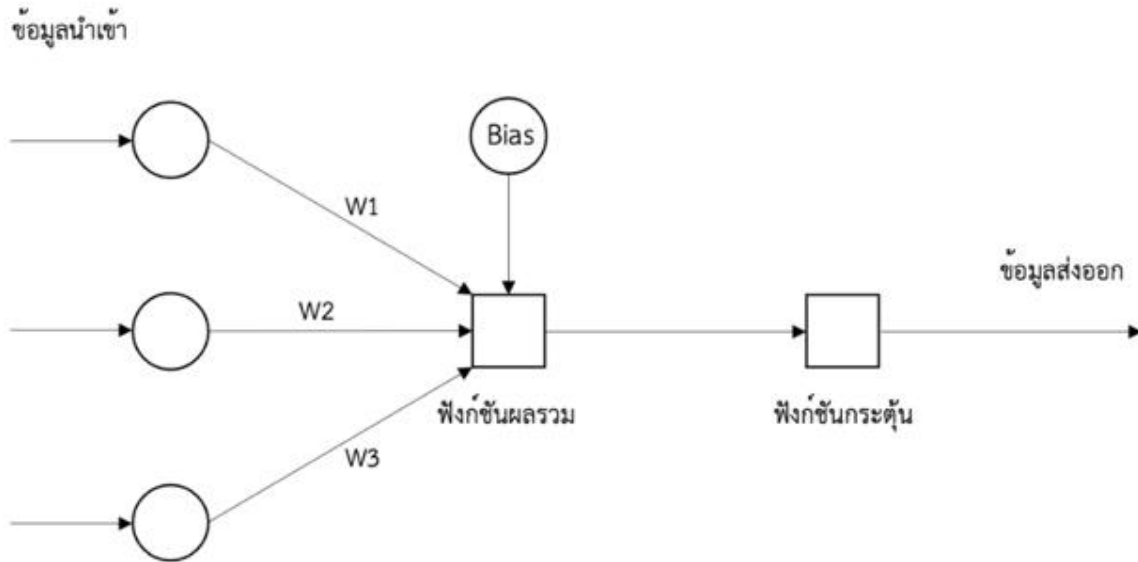
$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

4. โครงข่ายประสาทเทียม (Artificial Neural Network) การศึกษาในเรื่องนี้เริ่มต้นขึ้นในปี พ.ศ. 2486 เมื่อ McCulloch และ Pitts (Warren & Walter, 1943) ได้นำเสนอแบบจำลองทางคณิตศาสตร์ที่แสดงให้เห็นถึงแบบจำลองเซลล์ประสาทและพยายามทำความเข้าใจการทำงานของสมองร่วมกับกฎของ Hebb ซึ่งเสนอขึ้นในปี พ.ศ. 2492 ที่แสดงให้เห็นถึงกระบวนการเรียนรู้ที่เกิดขึ้นระหว่างเซลล์ที่เชื่อมต่อกัน (Hebb, 2002) ถือเป็นจุดเริ่มต้นของการศึกษาในเรื่องนี้ จากที่ได้กล่าวไปข้างต้นจะเห็นได้ว่าโครงข่ายประสาทเทียมมีต้นแบบแนวคิดมาจากการทำงานของสมอง

มนุษย์ ซึ่งมีส่วนย่อยคือเซลล์ประสาท (Neuron) เชื่อมต่อกันเป็นโครงข่าย บริเวณที่เชื่อมต่อกันเรียกว่าไซแนปส์ (Synapses) ซึ่งในปี พ.ศ. 2501 Rosenblatt (Rosenblatt, 1958) ได้นำเสนอถึงโครงข่ายประสาทเทียมแบบเพอร์เซปตรอน (Perceptron) ซึ่งถือเป็นรูปแบบพื้นฐานและง่ายที่สุดของโครงข่ายประสาทเทียม จากรูปที่ 4 จะแสดงให้เห็นถึงแบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอน ปมที่มีลักษณะวงกลมแสดงให้เห็นถึงเซลล์ประสาท แทนการเชื่อมต่อไซแนปส์ด้วยเส้นเชื่อมซึ่งถูกกำกับไว้ด้วยค่าถ่วงน้ำหนัก (Weight) ในแต่ละเส้น มีฟังก์ชันผลรวมและค่า

ความเอนเอียง (Bias) สำหรับการเชื่อมต่อสื่อสารของเซลล์ประสาทจะเป็นลักษณะของการกระตุ้นด้วยศักย์ไฟฟ้าหรือที่เรียกว่าฟังก์ชันกระตุ้น

(Activation Function) เมื่อการกระตุ้นถึงจุดที่กำหนดการส่งกระแสประสาทจะเกิดขึ้น (Aggarwal, 2018)



ภาพประกอบที่ 4 โครงข่ายประสาทเทียม

5. Gradient Boosting with XGBoost เป็นเทคนิคที่มีพื้นฐานมาจากการเรียนรู้แบบ Ensemble เช่นเดียวกับเทคนิคป่าแบบสุ่มกล่าวคือเป็นการเรียนรู้จากหลายโมเดลร่วมกัน แต่ Gradient Boosting จะเลือกใช้เทคนิคการเรียนรู้แบบทำซ้ำตามลำดับ ซึ่งจะเป็นการปรับปรุงโมเดลไปทีละขั้นในแต่ละรอบของการเรียนรู้ ส่งผลให้มีความแม่นยำที่สูงกว่าแต่ก็สามารถเกิดปัญหา Overfitting ได้ง่าย ในส่วนของ Extreme Gradient Boosting หรือ XGBoost เป็นเทคนิคที่ปรับปรุงมาจาก Gradient Boosting เพื่อให้สามารถทำงานได้อย่างรวดเร็ว และมีประสิทธิภาพมากยิ่งขึ้น รวมทั้งยังสามารถลดการเกิด Overfitting ได้ (Korstanje, Gradient Boosting with XGBoost and LightGBM, 2021)

การประเมินโมเดล

ขั้นตอนถัดมาหลังจากที่ได้ทำการสร้างโมเดลเป็นที่เรียบร้อยแล้วคือการประเมินโมเดล ซึ่งก็

คือ ประเมินความสามารถและความแม่นยำในการทำงานของโมเดล เพื่อทำการตัดสินใจว่าโมเดลที่ได้ทำการ สร้างขึ้นมานั้นเหมาะสมที่จะนำไปใช้หรือไม่ หรือควรทำการปรับแต่งให้มีความเหมาะสมก่อน เนื่องจาก ผู้วิจัยเลือกใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอนที่สามารถจำแนกได้เป็นสองประเภทคือ การ ถดถอยและการจำแนก ดังนั้นการประเมินโมเดลจึงมีความแตกต่างกันตามประเภทของการนำไปใช้ ในงานวิจัยฉบับนี้ต้องการ คำนวณการสูญเสียของชิ้นส่วนเครื่องจักรซึ่งเป็นลักษณะข้อมูลที่มี การอ้างอิงตามเวลาจึงจะใช้ อัลกอริทึมประเภทของการถดถอยและจะให้ค่า ผลลัพธ์ในลักษณะของค่า ตัวเลข ดังนั้นการ ประเมินโมเดลจะใช้การวัดระยะที่คลาดเคลื่อน (Error) ด้วยวิธีดังต่อไปนี้

1. สัมประสิทธิ์การกำหนด (Coefficient of Determination: R2)

$$R^2 = \frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma(y - \bar{y})^2}$$

2. ความผิดพลาดกำลังสองเฉลี่ย (Mean Square Error: MSE) และรากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error: RMSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_i)^2$$

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

เมื่อ y คือค่าจริง (Actual Value)

$\hat{y}$ คือค่าทำนาย (Predicted Value)

$\bar{y}$ คือค่าเฉลี่ยของค่าจริง

n คือจำนวนข้อมูล

การปรับปรุงโมเดล

หากพบว่าโมเดลยังไม่มีความแม่นยำ หรือไม่มีความเหมาะสมที่จะนำไปใช้จากขั้นตอนการประเมินโมเดลก็ต้องนำมาพิจารณาถึงสาเหตุและทำการปรับปรุงโมเดลให้มีความสมบูรณ์เพียงพอต่อการนำไปใช้งาน ปัญหาที่พบโดยส่วนใหญ่นี้คือปัญหา Overfitting ซึ่งจะส่งผลให้โมเดลนั้นมีประสิทธิภาพต่ำเมื่อนำไปใช้กับข้อมูลที่ไม่เคยพบ ดังนั้นจึงต้องทำให้โมเดลสามารถนำไปใช้ได้กับข้อมูลทั่วไปที่โมเดลไม่เคยพบได้อย่างมีประสิทธิภาพ หรือที่เรียกว่าการทำ Generalization

การเปรียบเทียบผลลัพธ์

เมื่อผ่านขั้นตอนของการปรับปรุงโมเดลให้มีประสิทธิภาพมากยิ่งขึ้นแล้ว โมเดลที่ถูกสร้างขึ้นจากห้าเทคนิคได้แก่ ต้นไม้ตัดสินใจ ป่าแบบสุ่ม ขั้นตอนวิธีเพื่อนบ้านใกล้สุด โครงข่ายประสาทเทียม และ Gradient Boosting with XGBoost จะถูกนำมาเปรียบเทียบประสิทธิภาพกันด้วยการ

วัดค่าสัมประสิทธิ์การกำหนด และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย โดยจะทำการแสดงผลในรูปแบบของตารางแสดงการเปรียบเทียบ

สรุปผลการวิจัย

งานวิจัยฉบับนี้ได้ทำการสร้างโมเดลจากห้าเทคนิคการเรียนรู้ของเครื่องได้แก่ ต้นไม้ตัดสินใจ ป่าแบบสุ่ม ขั้นตอนวิธีเพื่อนบ้านใกล้สุด โครงข่ายประสาทเทียม และ Gradient Boosting with XGBoost ในแต่ละโมเดลการเรียนรู้ของเครื่องที่ถูกนำมาใช้ในการเปรียบเทียบประสิทธิภาพนั้นจำเป็นต้องได้รับการฝึกฝนให้สามารถคาดการณ์ความเสียหายของชิ้นส่วนอากาศยานจากชุดข้อมูลที่เตรียมไว้ พร้อมทั้งเปรียบเทียบประสิทธิภาพในการคาดการณ์ความเสียหายของชิ้นส่วนอากาศยานโดยใช้ชุดข้อมูลจากโมเดล N-CMAPSS ซึ่งมีจำนวนข้อมูลทั้งหมด 688,351 ชุด 28 แอตทริบิวต์ แบ่งเป็น 552,088 และ 136,263 ชุดข้อมูลสำหรับฝึกฝนและทดสอบ

โมเดลตามลำดับ ค่าสัมประสิทธิ์การกำหนด (Coefficient of Determination: R2) และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root

Mean Square Error: RMSE) จะถูกนำมาใช้เพื่อประเมินประสิทธิภาพของโมเดล ซึ่งผลลัพธ์ที่ได้เปรียบเทียบกับไว้ดังตารางที่ 1

ตารางที่ 1 ประสิทธิภาพของโมเดลในแต่ละเทคนิค

Technique	R ²	RMSE
Decision Tree	- 0.2486	2.5157
Random Forest	0.4162	1.4057
kNN	0.4543	2.9364
Artificial Neural Network	0.7578	0.0016
Gradient Boosting with XGBoost	0.3175	0.0038

เมื่อทำการประเมินประสิทธิภาพแต่ละพารามิเตอร์ที่ใช้แสดงสถานการณ์เสื่อมสภาพ

(Model health parameters) ของโมเดลทั้ง 5 เทคนิค ได้ผลลัพธ์แสดงดังตารางที่ 2 ถึงตารางที่ 6

ตารางที่ 2 ประสิทธิภาพของพารามิเตอร์แสดงสถานการณ์เสื่อมสภาพจากเทคนิคต้นไม้ตัดสินใจ

Parameters	R ²	RMSE
fan_eff_mod	- 0.0540	0.0091
fan_flow_mod	0.0305	0.0066
LPC_eff_mod	- 0.5703	0.0042
LPC_flow_mod	- 3.1438	0.0041
HPC_eff_mod	0.5778	0.0018
HPC_flow_mod	0.6373	0.0032
HPT_eff_mod	0.5512	0.0020
HPT_flow_mod	- 1.4292	0.0016
LPT_eff_mod	0.0606	0.0029
LPT_flow_mod	0.0504	0.0036

ตารางที่ 3 ประสิทธิภาพของพารามิเตอร์แสดงสถานการณ์เสื่อมสภาพจากเทคนิคป่าแบบสุ่ม

Parameters	R ²	RMSE
fan_eff_mod	0.3864	0.0091
fan_flow_mod	0.4615	0.0066
LPC_eff_mod	0.1813	0.0042
LPC_flow_mod	- 0.5709	0.0040
HPC_eff_mod	0.8433	0.0018
HPC_flow_mod	0.8651	0.0032

HPT_eff_mod	0.7583	0.0020
HPT_flow_mod	0.3739	0.0016
LPT_eff_mod	0.5462	0.0029
LPT_flow_mod	0.4635	0.0036

ตารางที่ 4 ประสิทธิภาพของพารามิเตอร์แสดงสถานการณ์เสื่อมสภาพจากเทคนิคขั้นตอนวิธีเพื่อนบ้านใกล้สุด

Parameters	R ²	RMSE
fan_eff_mod	0.0607	0.0086
fan_flow_mod	0.1389	0.0062
LPC_eff_mod	- 0.0906	0.0035
LPC_flow_mod	- 1.1002	0.0029
HPC_eff_mod	0.4164	0.0022
HPC_flow_mod	0.5936	0.0033
HPT_eff_mod	0.5718	0.0020
HPT_flow_mod	- 0.3285	0.0012
LPT_eff_mod	0.2453	0.0026
LPT_flow_mod	0.2422	0.0032

ตารางที่ 5 ประสิทธิภาพของพารามิเตอร์แสดงสถานการณ์เสื่อมสภาพจากเทคนิคโครงข่ายประสาทเทียม

Parameters	R ²	RMSE
fan_eff_mod	0.8996	0.0028
fan_flow_mod	0.8861	0.0023
LPC_eff_mod	0.5776	0.0022
LPC_flow_mod	0.1710	0.0018
HPC_eff_mod	0.8961	0.0009
HPC_flow_mod	0.9103	0.0016
HPT_eff_mod	0.7860	0.0014
HPT_flow_mod	0.1307	0.0009
LPT_eff_mod	0.6811	0.0017
LPT_flow_mod	0.5697	0.0024

ตารางที่ 6 ประสิทธิภาพของพารามิเตอร์แสดงสถานะการเสื่อมสภาพจากเทคนิค Gradient Boosting with XGBoost

Parameters	R^2	RMSE
fan_eff_mod	0.1949	0.0079
fan_flow_mod	0.2285	0.0059
LPC_eff_mod	0.2019	0.0030
LPC_flow_mod	- 0.2051	0.0022
HPC_eff_mod	0.6950	0.0015
HPC_flow_mod	0.6597	0.0030
HPT_eff_mod	0.6761	0.0017
HPT_flow_mod	- 0.0057	0.0010
LPT_eff_mod	0.3812	0.0024
LPT_flow_mod	0.3488	0.0030

จากการประเมินประสิทธิภาพของโมเดลจากเทคนิคการเรียนรู้ของเครื่องทั้งหมด 5 เทคนิคพบว่าเทคนิคโครงข่ายประสาทเทียมให้ผลลัพธ์ที่ดีที่สุดด้วยค่าค่าสัมประสิทธิ์การกำหนด (R^2) และค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE) เท่ากับ 0.7578 และ 0.0016 ตามลำดับ เมื่อวิเคราะห์ถึงชุดข้อมูลที่น่าสนใจในงานวิจัยนี้จะพบว่ามีค่าความซับซ้อนสูงและไม่มีความเป็นเชิงเส้น (Non-Linear) จึงทำให้เทคนิคพื้นฐานมีความคลาดเคลื่อนในการทำนายสูง ในขณะที่เทคนิคโครงข่ายประสาทเทียมสามารถรองรับข้อมูลที่มีลักษณะดังกล่าวได้ดีกว่า

ข้อเสนอแนะ

งานวิจัยนี้ได้ทำการเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องสำหรับคาดการณ์ความเสียหายของเครื่องยนต์อากาศยาน นอกเหนือจากที่ทางผู้วิจัยได้ทดลองไป ยังมีเทคนิคอื่นๆ ที่น่าสนใจและสามารถนำมาใช้ในเชิงเดียวกันได้อาทิเช่น เทคนิคการเรียนรู้เชิงลึก เป็นต้น ซึ่งเทคนิคที่กล่าวมานั้นปัจจุบันได้มีการพัฒนาให้มีประสิทธิภาพสูงขึ้น จึงมีความน่าสนใจที่จะนำมาทดลองและเปรียบเทียบประสิทธิภาพต่อไป

เอกสารอ้างอิง

- มนัสกานต์ เสนหา. (2559). การทำนายอายุการใช้งานคงเหลือของเครื่องจักรด้วยเน็ตเวิร์กคอนโวลูชันเชิงลึกที่เพิ่มประสิทธิภาพ. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย.
- Junjie, L., Jinquan, H., & Feng, I. (1707). Distributed Kernel Extreme Learning Machines for Aircraft Engine Failure Diagnostics. Applied Sciences, 1-19.
- Efstratios , L. N., & Pantelis , B. (2015). Diagnostic methods for an aircraft engine performance. Journal of Engineering Science and Technology Review, 64-72.
- Khanh , T. N., & Kamal, M. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering & System Safety, 251-262.

- Gian, A. S., Andrea, S., Simone, P., & Sean, M. (2015). Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. *IEEE Transactions on Industrial Informatics*, 812-820.
- Ana, C., Jose, B., Paulo, L., Carla , A. G., Leonel, D., Jacinta, C., . . . Luís, R. (2018). Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. *IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, 4-7.
- Tiago, Z., Cristiano, A. C., Rodrigoda, R. R., Miromar, J. L., Eduardo , S. T., & Guann , P. L. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*.
- Thyago , P. C., Fabrizzio , A. S., Roberto, V., Roberto , d. F., João , P. B., & Symone , G. A. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*.
- Abhinav, S., Kai, G., Don, S., & Neil, E. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *International Conference on Prognostics and Health Management*.
- Manuel, A. C., Chetan, S. K., Kai, G., & Olga, F. (2021). Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics. *Data*, 1-14.
- Chao, C. K. (2022, January 26). NASA Ames Prognostics Data Repository. Retrieved from NASA Ames Prognostics Data Repository: <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
- J. R., Q. (1986). Induction of decision trees. *Machine Learning*, 81-106.
- Bahzad , T. J., & Adnan, M. A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS*, 20-28.
- Gaurav, S. (2020, June 25). Ensemble Methods in Machine Learning: Bagging Versus Boosting. Retrieved from PLURALSIGHT: <https://www.pluralsight.com/guides/ensemble-methods:-bagging-versus-boosting>
- Leo, B. (2001). Random Forests. *Machine Learning*, 5-32.
- Warren , S. M., & Walter , P. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 115-133.
- Hebb, D. (2002). *The Organization of Behavior*. New York: Psychology Press.
- Rosenblatt, F. (1958). THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN. *Psychological Review*, 386-408.
- Aggarwal, C. C. (2018). An Introduction to Neural Networks. In *Neural Networks and Deep Learning* (pp. 1-52). Switzerland: Springer.
- Korstanje, J. (2021). The kNN Model. In J. Korstanje, *Advanced Forecasting with Python With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR* (pp. 169-177). Maisons Alfort, France: Apress.

- Korstanje, J. (2021). Gradient Boosting with XGBoost and LightGBM. In J. Korstanje, Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR (pp. 193-205). New York: Apress.
- Vimala, M., Tom, T., Vikram, S., B, M., & M, G. (2017). Prediction of Remaining Useful Lifetime (RUL) of Turbofan Engine using Machine Learning. IEEE International Conference on Circuits and Systems (ICCS 2017), 306-311.
- Rounak, B., & Manikandan, J. (2021). Prediction of Remaining Useful Life for Aero-Engines. 2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES). Bali, Indonesia: IEEE .
- Nathaniel, D., Christopher, S., & Katherine, F. (2021). Inception based deep convolutional neural network for remaining useful life estimation of turbofan engines. International Journal of Prognostics and Health Management (pp. 1-6). PHM Society.