

การเปรียบเทียบประสิทธิภาพการแจกแจงไวบูล
และการตรวจจับความผิดปกติของเครื่องจักร
EFFICIENCY COMPARISONS OF MIXED WEIBULL DISTRIBUTION
AND MACHINE ANOMALY DETECTION



การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2566

KMITL-2023-SC-M-017-018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EFFICIENCY COMPARISONS OF MIXED WEIBULL DISTRIBUTION
AND MACHINE ANOMALY DETECTION



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
KMUTL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2023

KMITL-2023-SC-M-017-018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2023

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การเปรียบเทียบประสิทธิภาพการแจกแจงไวบูล และการตรวจจับความผิดปกติของเครื่องจักร
ชื่อนักศึกษา	นายณพล โรจนดำรงกุล
รหัสประจำตัว	64605035
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
พ.ศ.	ศุนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง 2566
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	รองศาสตราจารย์สายชล สินสมบูรณ์ทอง

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าความผิดปกติของเครื่องจักรจากข้อมูลการบำรุงรักษาและข้อมูลการใช้งานเครื่องจักร ในข้อมูลการบำรุงรักษาเครื่องจักรจะใช้การแจกแจงไวบูลจำนวน 4 วิธี คือ วิธีไวบูล วิธีไวบูลเชิงประกอบ วิธีไวบูลผสมเชิงเส้น และวิธีไวบูลผสมพหุนาม โดยพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบ ส่วนการใช้งานเครื่องจักรจะใช้การประมาณค่าความผิดปกติเป็น 2 รูปแบบคือ ความผิดปกติที่ไม่สัมพันธ์กับความเร็วของเครื่องจักร และความผิดปกติที่สัมพันธ์กับความเร็วของเครื่องจักร ความผิดปกติที่ไม่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดปกติจาก 2 วิธี คือ วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและวิธีป่าไม้โตดเดียว ส่วนความผิดปกติที่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดปกติจาก 3 วิธี คือ วิธีแผนภูมิการถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธี k เท่าของส่วนเหลือมาตรฐาน พิจารณาจากค่าความเที่ยง ค่าเรียกคืน และคะแนน $F1$

จากการศึกษาข้อมูลการบำรุงรักษาพบว่าวิธีไวบูลผสมพหุนามเป็นวิธีที่มีรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำที่สุด ส่วนข้อมูลการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วของเครื่องจักรพบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวเป็นวิธีที่มีค่าความเที่ยง ค่าเรียกคืน และคะแนน $F1$ สูงที่สุด และข้อมูลการใช้งานที่สัมพันธ์กับความเร็วของเครื่องจักรพบว่าวิธี k เท่าของส่วนเหลือมาตรฐานเป็นวิธีที่มีค่าความเที่ยง ค่าเรียกคืน และคะแนน $F1$ สูงที่สุด

คำสำคัญ : ไวบูล ไวบูลเชิงประกอบ ไวบูลผสมเชิงเส้น ไวบูลผสมพหุนาม k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร k เท่าของส่วนเหลือมาตรฐาน ป่าไม้โตดเดียว แผนภูมิการถดถอย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Efficiency comparisons of mixed Weibull distribution and machine anomaly detection
Student Name	Naphol Rochanadumrongkul
Student ID	64605035
Degree	Master of Science (Data science and analytics) KMITL-Digital Analytics and Intelligence Center
Year	2023
Independent Study Advisor	Assoc. Prof. Saichon Sinsomboonthong

Abstract

The objective of this research is to study and compare methods for estimating the anomaly score of machines from maintenance and operation data. In the maintenance data, four methods of Weibull distribution based are used: the traditional Weibull distribution, the composite distribution, the linear mixed distribution, and the polynomial mixed distribution. The root mean square error and mean absolute error is used as a criterion for comparison. In the operation data, two types of anomaly estimation are used: anomaly unrelated to machine speed and anomaly related to machine speed. For anomaly unrelated to machine speed, the estimation will be calculated using two methods: the k of single-variable standard deviation and isolation forests method. For anomaly related to machine speed, the estimation will be calculated using three methods: the regression control, the k of multiple-variable standard deviation and the k of standard residual. The two types of estimating anomaly will be tested by precision, recall and F1 score.

The study found that the polynomial mixed distribution is the method that has the lowest average root mean square error and average absolute error and the k of single-variable standard deviation is the method that has the highest precision, recall, and F1 score for anomaly unrelated to machine speed, while the k of standard residual is the method that has the highest precision, recall, and F1 score for anomaly related to machine speed.

Keywords : Weibull distribution, Composite distribution, Linear mixed distribution, Polynomial mixed distribution, K of single-variable standard deviation, K of multiple-variable standard deviation, K of standard residual, Isolation forests, Regression control



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพการแจกแจงผสมไวบูล และการตรวจจับความผิดปกติของเครื่องจักร สามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความอนุเคราะห์และสนับสนุนเป็นอย่างดีจาก รศ.สายชล สินสมบุญรณทอง อาจารย์ที่ปรึกษาการค้นคว้าอิสระที่ได้กรุณาให้คำปรึกษา ความรู้ ข้อคิด ข้อเสนอแนะ และปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งงานวิจัยครั้งนี้สำเร็จเรียบร้อยด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณส่วนบริหารการบำรุงรักษาสถานีเพิ่มความดันก๊าซ สายงานระบบท่อส่งก๊าซ ธรรมชาติ ที่ให้ความอนุเคราะห์ข้อมูลการบำรุงรักษาและข้อมูลการใช้งานเครื่องจักรที่ใช้ในงานวิจัยครั้งนี้

ขอขอบคุณอาจารย์และเพื่อนคณะวิทยาศาสตร์สาขาวิทยาการข้อมูลและการวิเคราะห์ทุกท่านที่ให้ความรู้และช่วยเหลือในการทำวิจัยครั้งนี้

ขอขอบคุณครอบครัวที่ให้ความอนุเคราะห์อนุมัติงบประมาณและอำนวยความสะดวกในการทำวิจัยครั้งนี้

สุดท้ายนี้ผู้วิจัยหวังว่างานวิจัยฉบับนี้คงเป็นประโยชน์สำหรับหน่วยงานที่เกี่ยวข้องและผู้ที่เกี่ยวข้องศึกษาต่อไป

ผู้วิจัย

นายณพล โรจนดำรงกุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ง
สารบัญ.....	จ
สารบัญตาราง.....	ช
สารบัญรูป.....	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	3
1.3 ขอบเขตของงานวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	6
2.1 ทฤษฎีที่เกี่ยวข้อง.....	6
2.2 งานวิจัยที่เกี่ยวข้อง.....	24
บทที่ 3 วิธีการดำเนินงานวิจัย	24
3.1 ข้อมูลการบำรุงรักษาเครื่องจักร (Maintenance Data).....	24
3.2 ข้อมูลการใช้งานเครื่องจักร (Operation Data).....	27
3.3 การหาค่าความผิดปกติร่วม.....	31
3.4 เครื่องมือที่ใช้ในงานวิจัย.....	32
3.5 สรุปขั้นตอนการดำเนินงาน.....	33
บทที่ 4 ผลการวิจัยและการอภิปรายผล	34
4.1 ผลการวิจัยข้อมูลการบำรุงรักษาเครื่องจักร.....	34
4.2 ผลการวิจัยข้อมูลการใช้งานเครื่องจักร.....	44
4.3 การหาค่าความผิดปกติร่วม.....	54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.4 การอภิปรายผล.....	55
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	55
5.1 สรุปผลการวิจัย.....	55
5.2 ข้อเสนอแนะ.....	55
เอกสารอ้างอิง	57
ภาคผนวก.....	62
ภาคผนวก ก.....	62
ภาคผนวก ข.....	73
ประวัติผู้เขียน.....	85



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 การแปลงข้อมูลด้วยวิธียกกำลังของข้อมูลด้วยค่าพารามิเตอร์.....	11
4.1 อายุการใช้งานเครื่องจักรของแต่ละรูปแบบความผิดปกติ.....	34
4.2 ความน่าจะเป็นสะสมของความผิดปกติในแต่ละรูปแบบ.....	35
4.3 พารามิเตอร์การแจกแจงไวบูลในแต่ละรูปแบบความผิดปกติ.....	36
4.4 ความน่าจะเป็นสะสมของความผิดปกติของเครื่องจักร.....	36
4.5 ความน่าจะเป็นสะสมของการแจกแจงไวบูลแต่ละรูปแบบความผิดปกติ.....	38
4.6 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นสะสมของความผิดปกติแต่ละรูปแบบ.....	39
4.7 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นความผิดปกติสะสมแต่ละรูปแบบความผิดปกติ	40
พหุนามอันดับ 2	
4.8 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นความผิดปกติสะสมแต่ละรูปแบบความผิดปกติ	41
พหุนามอันดับ 3	
4.9 ความน่าจะเป็นสะสมของความผิดปกติแต่ละวิธี.....	43
4.10 จำนวนตัวแบบการใช้งานเครื่องจักรในแต่ละประเภทของตัวแบบ.....	44
4.11 ตัวแบบการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร.....	45
4.12 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของตัวแบบการใช้งานเครื่องจักร	46
ที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร	
4.13 แสดงผลการเปรียบเทียบค่าความแม่นยำ ค่าความถูกต้อง และคะแนน F1.....	48
4.14 ผลการตรวจสอบความสัมพันธ์เชิงเส้นกับความเร็วรอบเครื่องจักรและ.....	49
การแจกแจงของส่วนเหลือ	
4.15 ขีดจำกัดควบคุมของตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับ.....	51
ความเร็วรอบของเครื่องจักรวิธีแผนภูมิควบคุมถดถอย	
4.16 ค่าเฉลี่ยและค่าความแปรปรวนร่วมผกผันกับความเร็วรอบของเครื่องจักร.....	52
4.17 ขีดจำกัดควบคุมของตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักร.....	52
วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐาน	
4.18 ตารางตัวอย่างแสดงผลการเปรียบเทียบค่าความแม่นยำ ค่าความถูกต้อง และคะแนน F1.....	53
4.19 ตารางแสดงค่าความผิดพลาดงานข้อมูลการใช้งานเครื่องจักรแต่ละตัวแบบ.....	54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 อัตราความผิดปกติรูปร่าง (Bathtub shape)	9
2.2 ส่วนที่เหลือจากเส้นตรงที่เหมาะสม	13
2.3 ตัวอย่างแผนภูมิควบคุมถดถอย.....	20
2.4 การแจกแจงความน่าจะเป็นสะสมของการแจกแจงปกติเทียบกับ..... ช่วงการเบี่ยงเบนจากค่าเฉลี่ยที่ยอมรับได้	21
3.1 แผนภูมิแสดงความสัมพันธ์ตัวแปรต้นและตัวแปรตาม	26
3.2 ขั้นตอนการทำงานเพื่อคำนวณค่าความผิดปกติของเครื่องจักร.....	33
4.1 แผนภูมิแสดงความน่าจะเป็นสะสมของความผิดปกติแต่ละรูปแบบ.....	37
4.2 ต้นไม้ตัดสินใจการแบ่งกลุ่มช่วงอายุการใช้งานทุกรูปแบบความผิดปกติ.....	37
4.3 แผนภูมิแสดงความผันแปรของส่วนที่เหลือที่ได้จากการแจกแจงไวบูลผสมเชิงเส้น.....	39
4.4 แผนภูมิแสดงความผันแปรของส่วนที่เหลือที่ได้จากการแจกแจงไวบูลผสม..... พหุนามอันดับ 2	40
4.5 แผนภูมิแสดงความผันแปรของส่วนที่เหลือที่ได้จากการแจกแจงไวบูลผสม..... พหุนามอันดับ 3	41
4.6 แผนภูมิแสดงความสัมพันธ์ของตัวแบบที่ได้จากข้อมูลการบำรุงรักษา.....	42
4.7 แผนภูมิแสดงค่าเฉลี่ยของความผิดพลาดสัมบูรณ์และค่ารากของค่าคลาดเคลื่อนกำลังสอง..... เฉลี่ยของการแจกแจงแต่ละวิธี	44
4.8 ข้อมูลแสดงการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบวิธีป่าไม้โตคนเดียว.....	47
4.9 ตัวอย่างข้อมูลทดสอบการใช้งานในสภาวะปกติและผิดปกติ.....	47
4.10 แสดงตัวอย่างข้อมูลการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักร.....	48
4.11 ตัวอย่างข้อมูลทดสอบการใช้งานในสภาวะปกติและผิดปกติ.....	53
4.12 ตัวอย่างการแจ้งเตือนค่าความผิดปกติมากกว่า 20%.....	54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องด้วยความต้องการใช้พลังงานในประเทศไทยมีแนวโน้มเพิ่มสูงขึ้น และในปัจจุบันได้มีการนำเข้าพลังงานจากต่างประเทศเพื่อให้เพียงพอต่อความต้องการ ก๊าซธรรมชาติเป็นแหล่งพลังงานหลักของภาคการผลิตไฟฟ้า ภาคอุตสาหกรรม และภาคยานยนต์ ดังนั้นในการสร้างความมั่นคงทางพลังงานของประเทศจำเป็นต้องส่งต่อก๊าซธรรมชาติจากแหล่งผลิตไปถึงผู้ใช้งานได้อย่างต่อเนื่องและปลอดภัย ระบบขนส่งก๊าซธรรมชาติเป็นวิธีหนึ่งในการขนส่งก๊าซธรรมชาติที่ได้รับความนิยม โดยมีสถานีเพิ่มความดันก๊าซธรรมชาติทำหน้าที่เพิ่มความดันก๊าซธรรมชาติให้ไปถึงผู้ใช้งานได้อย่างต่อเนื่องและปลอดภัย แต่ด้วยข้อจำกัดทางทรัพยากร เช่น งบประมาณ จำนวนบุคลากร และ อุปกรณ์ เป็นต้น จึงมีงานวิจัยจำนวนมากที่ต้องการทำนายความผิดปกติที่อาจเกิดขึ้น เพื่อเป็นแนวทางในการกำหนดแผนการบำรุงรักษาและการทำงานของเครื่องจักร ในงานวิจัยนี้จึงมีจุดประสงค์ในการหาค่าความผิดปกติของเครื่องจักรในสถานีเพิ่มความดันก๊าซธรรมชาติเพื่อลดความไม่พร้อมของเครื่องจักร ความไม่พร้อมของเครื่องจักรจะเกิดขึ้นเมื่อเครื่องจักรหยุดการผลิตโดยไม่ได้มีการวางแผนไว้ อาจเกิดจากการทำงานผิดปกติหรือมีความผิดปกติจากอุปกรณ์ส่วนย่อย

จากการศึกษาของ Kusiak and Verma (2012) ศึกษาความผิดปกติของกังหันลมจากข้อมูลการใช้งานทั้งหมด 18 สถานี สามารถแบ่งความผิดปกติได้เป็น 3 กลุ่มคือ ความผิดปกติทั่วไป, ความผิดปกติเฉพาะของระบบ และความผิดปกติที่คาดไม่ถึง โดยเปรียบเทียบผลการทำนายความผิดปกติจากวิธีการเรียนรู้ของเครื่องจำนวน 5 วิธีคือ วิธีซัพพอร์ตเวกเตอร์แมชชีน, วิธีต้นไม้ตัดสินใจ, วิธีโครงข่ายประสาทเทียม, วิธีต้นไม้ทวิภาค และ วิธีป่าสุ่มตัดสินใจ ผลการวิจัยระบุว่าวิธีป่าสุ่มตัดสินใจมีค่าเฉลี่ยเรขาคณิตของค่าความเที่ยงในการทำนายสูงที่สุดที่ 99.4% ตามด้วยวิธีต้นไม้ทวิภาคที่ 98.8% Wang and Yin (2019) ศึกษาเกี่ยวกับการผสมผสานการแจกแจงไวบูลร่วมกับการวิเคราะห์อนุกรมเวลาในการทำนายอัตราความผิดปกติของสถานีไฟฟ้าย่อยเทียบกับการแจกแจงไวบูลเพียงอย่างเดียว ทำการเปรียบเทียบประสิทธิภาพด้วยวิธีส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ย พบว่าการแจกแจงไวบูลร่วมกับการวิเคราะห์อนุกรมเวลานั้นมีค่าส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ยน้อยกว่าการแจกแจงไวบูลเพียงอย่างเดียวเฉลี่ย 17% อาจเนื่องมาจากการแจกแจงไวบูลเพียงอย่างเดียวไม่สามารถอธิบายรูปแบบข้อมูลความผิดปกติแบบสุ่มได้ Jimenez et al. (2020) ได้ศึกษาการประมาณการอายุการใช้งานที่เหลืออยู่จากเซ็นเซอร์บนข้อมูลจำลองของเครื่องยนต์เจ็ทจำนวน 3 ชุดข้อมูล โดยเปรียบเทียบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การใช้โมเดลเดี่ยวและโมเดลกลุ่มในการทำนายการบำรุงรักษาด้วยวิธีค่าคลาดเคลื่อนกำลังสองเฉลี่ย จากการศึกษาพบว่าค่าคลาดเคลื่อนกำลังสองเฉลี่ยของโมเดลกลุ่มมีค่าน้อยกว่าโมเดลเดี่ยวทั้ง 3 ชุด ข้อมูล Yenel and Jimenez (2020) ศึกษาอายุการใช้งานเครื่องจักรเพื่อคาดการณ์จำนวนความผิดปกติของเครื่องจักรที่จะเกิดขึ้นในช่วงเวลา 12 เดือน หลังจากระยะเวลารับประกัน 60 เดือน หรือ 100,000 ไมล์ แล้วแต่ว่าจะถึงอย่างใดก่อน ผู้วิจัยได้วิเคราะห์ความน่าจะเป็นร่วมของเหตุการณ์ความผิดปกติของเครื่องจักรร่วมกับระยะเวลาการใช้งานเครื่องจักร ผลการศึกษาพบว่าวิธีการแจกแจงความน่าจะเป็นร่วมมีส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ยเท่ากับ 0.949 หน่วยและค่าเอนเอียงเท่ากับ -0.32 หน่วย ซึ่งต่ำกว่าวิธีการแจกแจงความน่าจะเป็นของเหตุการณ์ความผิดปกติเพียงอย่างเดียว Alfeo et al. (2022) ได้ศึกษาข้อมูลการใช้งานของตลับลูกปืนในชุดแบบจำลองเครื่องจักรประกอบด้วยข้อมูลการสั่น, อุณหภูมิ, สิ่งรบกวนภายนอก และอัตราการกินเชื้อเพลิง ผู้วิจัยได้นำข้อมูลเหล่านี้มาวิเคราะห์คุณลักษณะจากการเข้ารหัสเพื่อแยกคุณลักษณะที่สำคัญตามอนุกรมเวลา และใช้ในการพัฒนาตัวแบบทำนายกลุ่มข้อมูลความผิดปกติเทียบกับการแยกความผิดปกติโดยผู้เชี่ยวชาญ โดยเปรียบเทียบจากคะแนน F1 ซึ่งพบว่าตัวแบบสามารถแยกความผิดปกติได้ด้วยคะแนน F1 เฉลี่ยมากกว่า 0.94 ซึ่งมีค่าใกล้เคียงกับการแยกความผิดปกติโดยมนุษย์ Roy et al. (2021) ศึกษาการแจกแจงความผิดปกติของเครื่องจักรกังหัน โดยเปรียบเทียบการแจกแจงแบบเดียว การแจกแจงเชิงประกอบ และการแจกแจงผสม ด้วยวิธีทดสอบภาวะสารถูโดยใช้ตัวสถิติทดสอบคอลโมโกรอฟ-สมิร์นอฟ พบว่าการแจกแจงเชิงประกอบได้ผลการทดสอบคอลโมโกรอฟ-สมิร์นอฟที่ดีที่สุดเท่ากับ 0.1 ตามด้วยการแจกแจงผสม Cetin and Tasgin (2021) ศึกษาการหาค่าความผิดปกติของตัวแปรสุ่มอิสระ โดยใช้ข้อมูลสังเคราะห์ 7 กลุ่ม จำนวน 3,000 ตัวอย่างจากฟังก์ชัน `make_blobs` ใน `sklearn` package รวมกับสังเคราะห์ข้อมูลผิดปกติจากการสุ่มจำนวน 100 ตัวอย่าง ในงานวิจัยได้เปรียบเทียบขั้นตอนวิธีในการตรวจจับความผิดปกติจำนวน 4 วิธี ได้แก่ k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร, วิธีป่าไม้โตนเดียว, วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว และ วิธีการจัดกลุ่มเฉลี่ย k กลุ่ม ทำการเปรียบเทียบประสิทธิภาพด้วยค่าเรียกคืน, ความแม่นยำ, ความไว และ คะแนน F1 ซึ่งวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปรได้ค่าสูงที่สุดเท่ากับ 0.79 0.99 1.00 และ 0.88 ตามลำดับ ตามด้วยวิธีป่าไม้โตนเดียว

ดังนั้นจึงทำให้ผู้วิจัยมีความสนใจวิเคราะห์ค่าความผิดปกติของเครื่องจักรกังหัน ซึ่งเป็นเครื่องจักรสำคัญในสถานีเพิ่มความดันก๊าซธรรมชาติจากการหาค่าความผิดปกติร่วมระหว่างข้อมูลการบำรุงรักษาเครื่องจักรและข้อมูลการใช้งานเครื่องจักร ข้อมูลการบำรุงรักษาเครื่องจักรจะคำนวณค่าความผิดปกติจากการแจกแจงความน่าจะเป็นของความผิดปกติที่มีประสิทธิภาพสูงที่สุด จากการเปรียบเทียบการแจกแจงทั้ง 4 วิธี ได้แก่ การแจกแจงไวบูล การแจกแจงไวบลูประกอบ การแจกแจงเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไวบูลผสมเชิงเส้น และวิธีที่นำเสนอคือการแจกแจงไวบูลผสมพหุนาม และในส่วนข้อมูลการใช้งานเครื่องจักรจะใช้การประมาณค่าความผิดพลาดเป็น 2 รูปแบบคือ ความผิดพลาดที่ไม่สัมพันธ์กันความเร็วของเครื่องจักร และความผิดพลาดที่สัมพันธ์กับความเร็วของเครื่องจักร ความผิดพลาดที่ไม่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดพลาดจาก 2 วิธี คือ วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและวิธีป่าไม้โตดเดียว ความผิดพลาดที่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดพลาดจาก 3 วิธี คือ วิธีแผนภูมิการถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธี k เท่าของส่วนเหลือมาตรฐาน การประมาณค่าความผิดพลาดทั้ง 2 รูปแบบ จะประมาณค่า z ของวิธีที่มีประสิทธิภาพสูงที่สุด ซึ่งสุดท้ายจะนำไปเป็นข้อมูลในการกำหนดแนวทางการบำรุงรักษาและใช้งานเครื่องจักรได้อย่างเหมาะสม

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพการทำนายความผิดพลาดการแจกแจงไวบูล การแจกแจงไวบูลประกอบ การแจกแจงไวบูลผสมเชิงเส้น และการแจกแจงไวบูลผสมพหุนาม ในข้อมูลการบำรุงรักษาเครื่องจักร

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพการตรวจจับความผิดพลาดที่ไม่สัมพันธ์กับความเร็วของเครื่องจักรด้วย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว วิธีป่าไม้โตดเดียว วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร

1.2.3 เพื่อเปรียบเทียบประสิทธิภาพการตรวจจับความผิดพลาดที่สัมพันธ์กับความเร็วของเครื่องจักรด้วย วิธีแผนภูมิควบคุมถดถอย และวิธี k เท่าของส่วนเหลือมาตรฐาน ในข้อมูลการใช้งานเครื่องจักร

1.3 ขอบเขตของงานวิจัย

ในงานวิจัยนี้จะพิจารณาเฉพาะเครื่องจักรกังหันซึ่งเป็นเครื่องจักรหลักในสถานีเพิ่มความดันก๊าซธรรมชาติโดยใช้ข้อมูลในช่วงปี 2020 ถึงปี 2022 จำนวน 2 ส่วน ประกอบด้วยข้อมูลการบำรุงรักษาเครื่องจักรและข้อมูลการใช้งานเครื่องจักร โดยรวบรวมข้อมูลการบำรุงรักษาเครื่องจักร 119 ตัวอย่าง นำมาวิเคราะห์การแจกแจงความน่าจะเป็นของค่าความผิดพลาดจากการการแจกแจงไวบูล การแจกแจงไวบูลประกอบ การแจกแจงไวบูลผสมเชิงเส้น และการแจกแจงไวบูลผสมพหุนาม ด้วยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ส่วนข้อมูลการใช้งานเครื่องจักรจำนวน 100 ตัวแบบ จำนวนอย่างน้อย 7 วัน วันละ 24 ชั่วโมง หรือเท่ากับ 168 ตัวอย่าง นำมาวิเคราะห์ค่าความผิดพลาด ซึ่งแบ่งข้อมูลเป็น 2 รูปแบบคือ ข้อมูลที่ไม่ขึ้นกับความเร็วรอบของเครื่องจักรและข้อมูลที่ขึ้นกับความเร็วรอบของเครื่องจักร โดยที่ข้อมูลที่ไม่ขึ้นกับความเร็วรอบของเครื่องจักรจะเปรียบเทียบประสิทธิภาพการประมาณค่าความผิดพลาดจากวิธี k เท่าของส่วนเบี่ยงเบน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มาตรฐานตัวแปรเดียว และวิธีป่าไม้โตเตี่ยว ด้วยค่าค่าความเที่ยง ค่าเรียกคืน และ คะแนน F1 ส่วนข้อมูลที่ขึ้นกับความเร็รรอบของเครื่องจักรจะเปรียบเทียบประสิทธิภาพการการประมาณค่าความผิดพลาดจากวิธีแผนภูมิควบคุมการถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธีที่นำเสนอคือวิธี k เท่าของส่วนเหลือมาตรฐาน ด้วยค่าความเที่ยง ค่าเรียกคืน และคะแนน F1 จากนั้นนำวิธีที่มีประสิทธิภาพสูงสุดมาคำนวณค่าความผิดพลาดร่วมระหว่างข้อมูลการบำรุงรักษาเครื่องจักร และข้อมูลการใช้งานเครื่องจักร

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เพื่อเป็นแนวทางในการศึกษาการทำนายความผิดพลาดจากการแจจแจงความน่าจะเป็นของความผิดพลาดจากข้อมูลการบำรุงรักษาเครื่องจักร

1.4.2 เพื่อเป็นแนวทางในการศึกษาวิธีการหาค่าความผิดพลาดจากข้อมูลการใช้งานเครื่องจักร

1.4.3 เพื่อทราบถึงวิธีการคำนวณค่าความผิดพลาดของเครื่องจักรเพื่อเป็นแนวทางในการตัดสินใจปรับปรุงแผนการบำรุงรักษาให้เหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้จะกล่าวถึงทฤษฎีและแนวความคิดต่าง ๆ เอกสาร และผลการวิจัยที่เกี่ยวข้องกับการทำนายความผิดปกติของเครื่องจักร โดยนำมาบูรณาการให้เหมาะสมกับงานวิจัยในครั้งนี้ สามารถแบ่งหัวข้อสำคัญได้ดังนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ค่าใช้จ่ายที่เกิดขึ้นในสถานีเพิ่มความดันก๊าซธรรมชาติ (Kurz, 2012)

สถานีเพิ่มความดันก๊าซธรรมชาติมีส่วนสำคัญในการส่งผ่านก๊าซธรรมชาติให้กับภาคผลิตไฟฟ้า ภาคอุตสาหกรรม และภาคยานยนต์ เพื่อความมั่นคงทางด้านพลังงานของประเทศ แต่ด้วยข้อจำกัดทางด้านงบประมาณ ค่าใช้จ่ายที่เกิดขึ้นจึงจำเป็นต้องมีการบริหารที่เหมาะสม สามารถจำแนกค่าใช้จ่ายสำคัญ ได้แก่ ต้นทุนออกแบบและติดตั้ง ต้นทุนการดำเนินงาน (โดยเฉพาะต้นทุนเชื้อเพลิง) ต้นทุนการบำรุงรักษา และการปล่อยมลพิษ

นอกจากนี้การตัดสินใจเกี่ยวกับการออกแบบสถานีเพิ่มความดันก๊าซธรรมชาติ เช่น จำนวนเครื่องจักร จำนวนเครื่องจักรสำรอง ประเภทของเครื่องจักรขับเคลื่อน และประเภทของคอมเพรสเซอร์ มีผลกระทบต่อต้นทุน การสิ้นเปลืองเชื้อเพลิง ความยืดหยุ่นในการทำงาน การปล่อยมลพิษ และความพร้อมใช้งานของสถานีทั้งสิ้น ในงานวิจัยนี้มีจุดประสงค์เพื่อลดค่าใช้จ่ายในการบำรุงรักษาเท่านั้น ซึ่งค่าใช้จ่ายดังกล่าวหมายถึงค่าอะไหล่และค่าแรงเพื่อให้เครื่องจักรทำงานได้ตามที่ออกแบบไว้

2.1.2 รูปแบบการบำรุงรักษาร่วมสมัย (Kurz, 2012)

การบำรุงรักษาร่วมสมัยคือการกำหนดแผนการบำรุงรักษาตามรอบเวลาหรือตามเงื่อนไขการใช้งานของเครื่องจักร การบำรุงรักษาจะส่งผลต่อความพร้อมใช้งานของเครื่องจักรใน 2 วิธีคือ การบำรุงรักษาเนื่องจากเครื่องจักรเสียหาย การบำรุงรักษาประเภทนี้จะส่งผลให้ค่าความพร้อมใช้งานของเครื่องจักรลดลง ส่วนการบำรุงรักษาตามแผน การบำรุงรักษาประเภทนี้มักจะมีผลต่อความพร้อมใช้งานของเครื่องจักรน้อยกว่าการบำรุงรักษาที่ไม่ได้วางแผนไว้ เช่น การซ่อมใหญ่ (Overhaul) ตามกำหนดเวลาอาจทำให้อุปกรณ์ไม่สามารถใช้งานได้เพียงไม่กี่วันหากมีโครงการแลกเปลี่ยนเครื่องจักร

ในทางกลับกันการกำหนดแผนบำรุงรักษาที่ไม่เพียงพอหรือไม่เหมาะสมส่งผลเสียต่อความพร้อมใช้งาน เนื่องจากประสิทธิภาพการทำงานลดลงและมีโอกาสสูงที่เครื่องจักรจะไม่พร้อมใช้งานโดยไม่ได้วางแผนไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3 แนวทางการบำรุงรักษาในอุตสาหกรรม 4.0 (Matgorzata et al., 2019)

เนื่องด้วยอุตสาหกรรมกำลังประสบกับการเปลี่ยนแปลงภายใต้ความท้าทายของเศรษฐกิจและทรัพยากร ทางผู้ประกอบการจึงหาเทคโนโลยีขั้นสูงเพื่อเพิ่มขีดความสามารถในการแข่งขัน ในขณะที่เดียวกันการบำรุงรักษาที่กำลังเปลี่ยนบทบาทเพื่อรองรับการสร้างมูลค่าที่มากขึ้น ทั้งยังมีส่วนช่วยในด้านเศรษฐกิจและสิ่งแวดล้อม แนวคิดการบำรุงรักษา 4.0 จึงเกิดขึ้นมาเพื่อทดแทนการบำรุงรักษาแบบเก่าโดยทำให้ผู้ประกอบการสามารถเพิ่มอายุการใช้งานของเครื่องจักรในการผลิต ลดความไม่พร้อมของเครื่องจักรโดยไม่ได้วางแผนไว้ ลดการใช้พลังงานและทรัพยากร

เทคโนโลยีอุตสาหกรรม 4.0 ได้นิยามการบำรุงรักษาในรูปแบบการเรียนรู้ด้วยตนเอง มีระบบอัจฉริยะที่จะทำนายความผิดปกติ ทำให้เกิดการแจ้งเตือน ตรวจสอบ และกำหนดมาตรการป้องกันที่มีประสิทธิภาพโดยใช้เทคนิคการวิเคราะห์ขั้นสูงกับข้อมูลขนาดใหญ่ ซึ่งข้อมูลดังกล่าวประกอบด้วยข้อมูลทางเทคนิค ประวัติการบำรุงรักษา การใช้งาน สภาพแวดล้อม ประสิทธิภาพของเครื่องจักร และข้อมูลเครื่องจักรที่คล้ายคลึงกันที่อื่นๆ เป็นต้น

2.1.4 การตรวจสอบข้อมูลสำหรับพัฒนาต้นแบบในการวิจัย (Douglas and George, 2016)

การวิเคราะห์ข้อมูลในงานวิจัยนี้ใช้หลักการแจกแจงความน่าจะเป็น การแจกแจงความน่าจะเป็นคือตัวแบบทางคณิตศาสตร์ที่ใช้ในการอธิบายความสัมพันธ์ของตัวแปรสุ่มและโอกาสที่จะเกิดค่าตัวแปรสุ่มนั้นๆ ในประชากร สามารถแบ่งได้ 2 แบบคือ

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง จะถูกกำหนดโดยความสัมพันธ์ระหว่างค่าของตัวแปรสุ่มกับโอกาสที่จะเกิดค่าตัวแปรสุ่มนั้น

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง จะถูกกำหนดโดยความสัมพันธ์ระหว่างช่วงค่าของตัวแปรสุ่มกับโอกาสที่จะเกิดช่วงค่าตัวแปรสุ่มนั้น

ตัวแปรสุ่มที่กล่าวถึงนั้นคือค่าหรือลักษณะต่าง ๆ ที่สนใจเป็นค่าผลลัพธ์ที่ไม่คงที่เนื่องจากในระบบมีกลไกการเปลี่ยนแปลงที่อยู่นอกเหนือการควบคุม หรือเกิดจากการทดลองสุ่ม โดยผลลัพธ์ที่ได้ทั้งหมดจากการทดลองนั้นประกอบเป็นกลุ่มตัวอย่างของประชากรหรือระบบที่สนใจศึกษา ข้อมูลที่ใช้ในงานวิจัยจะเป็นตัวแปรสุ่มต่อเนื่องหรือตัวแปรสุ่มที่เมื่อทำการพิจารณาช่วงใด ๆ จำนวนตัวแปรสุ่มไม่สามารถระบุได้เป็นค่าคงที่ เนื่องจากตัวแปรสุ่มประเภทนี้ไม่สามารถนับแจกแจงสมาชิกได้ เพราะจำนวนสมาชิกไม่จำกัด ในการหาค่าความน่าจะเป็นจึงไม่สามารถหาได้จากสูตรทั่วไป ค่าความน่าจะเป็นจะทำการคำนวณภายในช่วงที่สนใจโดยใช้หลักการการหาพื้นที่เมื่อทราบฟังก์ชันความน่าจะเป็น แทนการหาสัดส่วนสมาชิกที่สนใจต่อสมาชิกในกลุ่มตัวอย่าง ฟังก์ชันดังกล่าวเรียกว่าฟังก์ชันการแจกแจงความน่าจะเป็นสำหรับตัวแปรสุ่ม ถ้าตัวแปรสุ่มต่อเนื่อง X มีค่าที่เป็นไปได้ภายในช่วง (a, b) ฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง (Probability Density Function) หรือ $f(x)$ จะมีสมบัติคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ความน่าจะเป็นมีค่าอยู่ในช่วง 0-1 เสมอ
2. ความน่าจะเป็น ณ จุดใดๆ หาค่าไม่ได้
3. พื้นที่รวมของค่าความน่าจะเป็นทุกค่าที่เป็นไปได้ของตัวแปรสุ่มมีค่าเท่ากับ 1

จากสมบัติดังกล่าวจึงได้มีการพิจารณาฟังก์ชันการแจกแจงความน่าจะเป็นสะสม (Cumulative Probability Distribution Function) หรือ $F(x)$ เพื่อระบุผลรวมของค่าความน่าจะเป็นจากค่าต่ำสุดที่เป็นไปได้ของตัวแปรสุ่มสะสมจนถึงค่า x ที่สนใจ

$$F(x) = \int_{-\infty}^x f_x(t) dt$$

ฟังก์ชันการแจกแจงความน่าจะเป็นสะสมมีสมบัติคือ

1. $0 \leq F(x) \leq 1$
2. $F(x) \leq F(y), X \leq Y$
3. $F(-\infty) = 0, F(\infty) = 1$
4. $f(x) = \frac{dF(x)}{dx}$
5. $P[a < x < b] = \int_a^b f(x)dx = F(b) - F(a)$

การวิเคราะห์ข้อมูลในงานวิจัยนี้จะพิจารณาข้อมูล 2 ส่วน ประกอบด้วยข้อมูลประวัติการบำรุงรักษาเครื่องจักร และข้อมูลการใช้งานเครื่องจักร ที่จะกล่าวถึงในข้อ 2.1.5 และ 2.1.6 ตามลำดับ จึงจำเป็นต้องคำนวณการแจกแจงความน่าจะเป็นรวมของตัวแปรสุ่มทั้ง 2 ส่วน ซึ่งการแจกแจงความน่าจะเป็นรวมมีคุณสมบัติดังนี้

1. $f(x, y) > 0$ สำหรับทุกค่าของ (x, y)
2. $\iint_{\forall x \forall y} f(x, y) dx dy = 1$
3. $P(X = x, Y = y) = 0$ ความน่าจะเป็น ณ จุดใดๆ หาค่าไม่ได้
4. $P((x, y) \in A) = \int_{\forall x \forall y \in A} f(x, y) dx dy$
5. $P(X \leq x, Y \leq y) = F(x, y) = \int_{\forall x \forall y \in A} f(x, y) dx dy$

หากทั้ง 2 เหตุการณ์ (x, y) มีความเป็นอิสระต่อกันสามารถหาการแจกแจงความน่าจะเป็นร่วมได้จาก

$$P(X \leq x, Y \leq y) = F(x, y) = F(x) \times F(y)$$

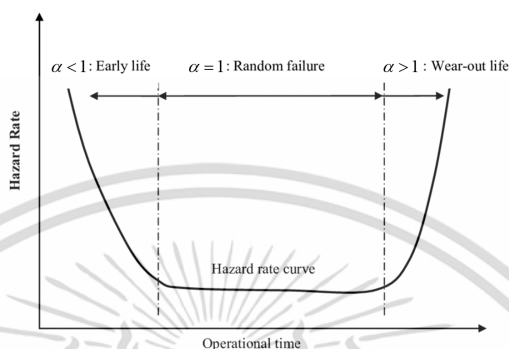
2.1.5 การวิเคราะห์ข้อมูลการบำรุงรักษาเครื่องจักร (Jiang, 2014)

ในงานวิจัยนี้จะใช้การแจกแจงไวบูลซึ่งเป็นที่ยอมรับใช้อย่างแพร่หลายในด้านแบบจำลองความผิดปกติของอุปกรณ์ การแจกแจงไวบูลเป็นการแจกแจงของความน่าจะเป็นต่อเนื่อง มีรูปมาตรฐานของการแจกแจงฟังก์ชันความหนาแน่นความน่าจะเป็นของ 2 ตัวแปรดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$f(t; \alpha, \beta) = \left(\frac{\alpha}{\beta}\right) \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}, t > 0$$

โดยที่ α เป็นตัวแปรรูปร่าง และ β เป็นตัวแปรบ่งขนาด หากพิจารณาอัตราความผิดพลาดตามช่วงอายุการใช้งานของเครื่องจักรหรืออัตราความผิดพลาดรูปร่าง (Bathtub shape) จะเห็นว่ามีการเปลี่ยนแปลงของค่าตัวแปรรูปร่างดังรูปที่ 2.1



รูปที่ 2.1 อัตราความผิดพลาดรูปร่าง

จากรูปที่ 2.1 ตัวแปรรูปร่างจะสามารถอธิบายอัตราความผิดพลาดได้ 3 ช่วงดังนี้ $\alpha < 1$ หมายถึงอัตราความผิดพลาดลดลงตามช่วงเวลา เช่น ความผิดพลาดที่เกิดจากกระบวนการผลิต

$\alpha = 1$ หมายถึงอัตราความผิดพลาดคงที่ตามช่วงเวลามีลักษณะสุ่ม เช่น ความผิดพลาดจากการใช้งานทั่วไป

$\alpha > 1$ หมายถึงอัตราความผิดพลาดเพิ่มขึ้นตามช่วงเวลา เช่น ความผิดพลาดที่เกิดจากการเสื่อมสภาพของอุปกรณ์

ในการวิเคราะห์ค่าความผิดพลาด สามารถพิจารณาจากฟังก์ชันความน่าจะเป็นสะสมดังนี้

$$F(t; \alpha, \beta) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha}$$

การประมาณค่าฟังก์ชันความน่าจะเป็นสะสมของการแจกแจงไวบูลจะประมาณด้วยค่าความผิดพลาดจากการเรียงลำดับช่วงเวลาก่อนเกิดความผิดพลาด (Time to fail) จากน้อยไปมาก และกำหนดความน่าจะเป็นสะสมของแต่ละลำดับ ซึ่งสามารถทำได้หลายวิธี เช่น ตัวเลขลำดับค่าเฉลี่ย อันดับมัธยฐาน การประมาณค่าของบิโนมรด์ การประมาณค่าของจอห์นสัน

งานวิจัยนี้ได้ใช้การประมาณค่าของบิโนมรด์ในการกำหนดค่าฟังก์ชันความหนาแน่นความน่าจะเป็นสะสมดังนี้ (Fernández and Vázquez 2012)

$$F_i(t) = \frac{i - 0.3}{n + 0.4}$$

โดยที่ i คือลำดับช่วงเวลาก่อนเกิดความผิดพลาด และ n คือขนาดตัวอย่างทั้งหมด เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น เมื่อผู้ใดเห็นใบใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.6 การวิเคราะห์ข้อมูลการใช้งานเครื่องจักร

ในงานวิจัยนี้จะมุ่งเน้นการตรวจจับความผิดปกติด้วยวิธีต่างๆ ในการตรวจสอบข้อมูลการใช้งานเครื่องจักรในสภาวะปกติที่นั่นส่วนใหญ่จะมีการแจกแจงปกติ ซึ่งการแจกแจงปกติจะเป็นการแจกแจงของตัวแปรสุ่มต่อเนื่องที่มีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(X | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

โดยที่ μ เป็นค่าเฉลี่ยของข้อมูล σ เป็นค่าส่วนเบี่ยงเบนมาตรฐาน และ X เป็นตัวแปรสุ่มของข้อมูลการใช้งานเครื่องจักรในแต่ละสัญญาณที่บันทึกได้ (Das, 2016) ในการตรวจสอบการแจกแจงปกติ นอกจากพิจารณากราฟของฟังก์ชันความหนาแน่นความน่าจะเป็นแล้ว ยังสามารถพิจารณาจากพล็อตควอนไทล์-ควอนไทล์ (Quantile-quantile plot) หรือ การทดสอบจาร์คูบีร่า (Jarque-Bera test) ซึ่งเป็นการทดสอบภาวะสารูปดี (Goodness-of-fit test) เป็นต้น

พล็อตควอนไทล์-ควอนไทล์ คือการเปรียบเทียบควอนไทล์ของการแจกแจงข้อมูลกับควอนไทล์ของการแจกแจงปกติทางทฤษฎี เมื่อพบควอนไทล์ของการแจกแจงที่ต้องการเปรียบเทียบหรือมีการแจกแจงตรงกับแจกแจงปกติ จุดจะอยู่ในระนาบใกล้เคียงเส้นตรง $y = x$ แต่หากแสดงจุดในลักษณะของเส้นโค้งที่มีความชันเพิ่มขึ้นจากซ้ายไปขวา แสดงว่าการแจกแจงของข้อมูลเอียงไปทางขวา และความชันลดลงจากซ้ายไปขวา แสดงว่าการแจกแจงข้อมูลเอียงไปทางซ้าย

การทดสอบจาร์คูบีร่าเดิมมีการเสนอโดย Bowman and Shenton (1975) เป็นการรวมกำลังสองของความเบ้และความโด่งในค่าสถิติเดี่ยวดังนี้

$$JB = \left(\frac{n}{6}\right) \times \left(S^2 + \frac{(K-3)^2}{4}\right)$$

$$\text{เมื่อ } S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{\frac{3}{2}}}$$

$$\text{และ } K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

โดยที่ \bar{x} เป็นค่าเฉลี่ยของข้อมูลตัวอย่าง หากข้อมูลมาจากการแจกแจงปกติ มีสมมุติฐานว่าง (Null hypothesis) เป็นสมมุติฐานร่วมของความเบ้ (S) ที่คาดหวังเท่ากับ 0 และความโด่ง (K) มากกว่า 0 ตามคำจำกัดความของ JB ค่าเบี่ยงเบนใด ๆ จากสมมุติฐานว่างจะทำให้ค่าสถิติ JB มีค่าเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.7 การแปลงข้อมูล

หากการตรวจสอบข้อมูลการใช้งานเครื่องจักรในสภาวะปกติไม่มีการแจกแจงปกติ เนื่องจากเงื่อนไขการใช้งาน งานวิจัยนี้จะแปลงข้อมูลด้วยวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ (Zhang et al., 2019) การแปลงแบบบ็อกซ์-ค็อกซ์นั้นนิยมใช้ในการปรับปรุงข้อมูลให้มีการแจกแจงปกติและมีการกระจายตัวคงที่ เป็นการแปลงข้อมูลด้วยวิธียกกำลังของข้อมูลด้วยค่าพารามิเตอร์

ตารางที่ 2.1 การแปลงข้อมูลด้วยวิธียกกำลังของข้อมูลด้วยค่าพารามิเตอร์

λ	$Y(\lambda)$
-3	$Y^{-3} = 1/Y^3$
-2	$Y^{-2} = 1/Y^2$
-1	$Y^{-1} = 1/Y^1$
-0.5	$Y^{-0.5} = 1/(\sqrt{Y})$
0	$\log(Y)$
0.5	$Y^{0.5} = \sqrt{Y}$
1	$Y^1 = Y$
2	Y^2
3	Y^3

จากตารางที่ 2.1 จะเห็นว่า $Y(\lambda)$ ขึ้นอยู่กับค่าสัมประสิทธิ์ของบ็อกซ์-ค็อกซ์ หากค่าสัมประสิทธิ์ของบ็อกซ์-ค็อกซ์ มีการเปลี่ยนแปลงค่า $Y(\lambda)$ จะเปลี่ยนตามแบบพหุนาม จำเป็นต้องตรวจสอบการแจกแจงปกติอีกครั้ง โดยที่

Y คือ ตัวแปรสุ่ม

λ คือ ค่าสัมประสิทธิ์ของบ็อกซ์-ค็อกซ์ สามารถหาค่าที่เหมาะสมได้จากสมมุติฐานที่ว่า $Y(\lambda)$ มีการแจกแจงปกติ

2.1.8 การพัฒนาตัวแบบในข้อมูลการบำรุงรักษาเครื่องจักร

2.1.8.1 ตัวแบบการถดถอยเชิงเส้น (Linear Regression Model)

การถดถอยเชิงเส้นเป็นการหาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามที่สอดคล้องตามสมมุติฐานหลัก 4 ข้อดังนี้ (Candance and Ronald, 2019)

1. ความเชื่อมโยงเชิงเส้น คือ ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามจะเป็นเชิงเส้นเมื่อการเปลี่ยนแปลงของตัวแปรตามจะสัมพันธ์ต่อการเปลี่ยนแปลงของตัวแปรอิสระ สามารถตรวจสอบได้จากค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเข้าใกล้ 1 หรือ -1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ความอิสระต่อกัน คือ ข้อมูลตัวแปรอิสระในชุดข้อมูลจะเป็นอิสระต่อกัน เมื่อไม่มีความสัมพันธ์ต่อตัวแปรอิสระตัวอื่นและไม่มีความสัมพันธ์ต่อส่วนเหลือ สามารถตรวจสอบได้จากการทดสอบเดออร์บินวัตสันที่มีค่าเข้าใกล้ 2
3. ความแปรปรวนคงที่ คือ ความแปรปรวนของส่วนเหลือควรมีค่าคงที่ในระดับต่าง ๆ ของตัวแปรอิสระในช่วงของค่าที่คาดการณ์ได้ สามารถตรวจสอบได้จากการทดสอบไวท์ที่สมมุติฐานว่างคือความแปรปรวนของส่วนเหลือควรมีค่าคงที่ (Homoscedasticity)
4. ความเป็นปรกติ คือ การแจกแจงของส่วนเหลือควรมีการแจกแจงปรกติ สามารถทดสอบได้จากการทดสอบจาร์คูบีว่าที่สมมุติฐานว่างคือข้อมูลมีการแจกแจงปรกติ

ส่วนการประมาณค่าพารามิเตอร์ของการถดถอยของประชากร ซึ่งการประมาณค่าพารามิเตอร์ที่นิยมนำมาใช้กันมากที่สุดคือ วิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Square) วิธีกำลังสองน้อยสุดสามัญมีแนวคิดคือ ถ้าเส้นสมการในความเป็นจริง (Real Line) คือ (Douglas et al., 2013)

$$E(Y) = \beta_0 + \beta_1 X$$

สมการดังกล่าวจะสามารถใช้แทนสมการที่แสดงความสัมพันธ์ได้เป็น

$$E(Y) = \beta_0 + \beta_1 X + \varepsilon_i$$

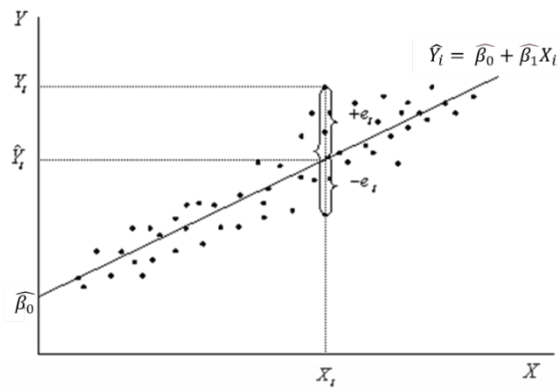
โดยต้นแบบความสัมพันธ์จะดีก็ต่อเมื่อผลรวมของความคลาดเคลื่อน $\sum \varepsilon_i$ มีค่าต่ำที่สุด แต่เนื่องจาก ε_i เป็นตัวแปรสุ่มที่วัดค่าไม่ได้ ดังนั้นจึงพิจารณาจากผลรวมกำลังสองของค่าประมาณของความคลาดเคลื่อน $\sum e_i^2$ มีค่าต่ำที่สุดแทน จะได้ความสัมพันธ์เป็น

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

$$Y_i = \hat{Y}_i + e_i$$

โดยที่ $\hat{\beta}_0, \hat{\beta}_1$ คือพารามิเตอร์ค่าประมาณที่อาศัยข้อมูลตัวอย่าง และเป็นตัวประมาณค่าของ β_0, β_1 และ \hat{Y}_i คือค่าที่ประมาณได้ของ Y_i และ e_i คือความคลาดเคลื่อน (Error) หรือส่วนที่เหลือ (Residual) สามารถแสดงในรูปส่วนเหลือได้

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \text{ โดยที่ } i = 1, 2, \dots, n$$



รูป 2.2 ส่วนที่เหลือจากเส้นตรงที่เหมาะสม

จากรูปที่ 2.2 มีข้อสังเกต 3 ประการคือ

1) การพิจารณาว่าเส้นสมการประมาณค่าใดเป็นเส้นสมการที่เหมาะสมที่สุด จะพิจารณาจากระยะห่างระหว่างค่าสังเกต (สนใจเฉพาะค่า Y_i) จากกลุ่มตัวอย่างกับเส้นสมการประมาณค่าที่สมมุติว่าเหมาะสมกับข้อมูลมากที่สุด ไม่ใช่ระยะห่างระหว่างค่าสังเกตจากกลุ่มตัวอย่างกับเส้นสมการจริง ที่เป็นเช่นนี้เพราะเส้นสมการจริงเป็นเส้นสมการในอุดมคติซึ่งไม่สามารถทราบตำแหน่งที่แท้จริงได้

2) ระยะห่างระหว่างค่าสังเกต Y_i กับค่าประมาณจากสมการ \hat{Y}_i คือระยะทางที่วัดจากจุดพิกัดของค่าสังเกตมายังเส้นสมการประมาณค่าตามแนวตั้งฉากกับแกน X หรือ $e_i = Y_i - \hat{Y}_i$ เมื่อ $i = 1, 2, \dots, n$ ถ้าจุดพิกัดอยู่เหนือเส้นสมการที่ประมาณได้จะถือว่า e_i มีค่าเป็นบวก และถ้าจุดพิกัดอยู่ใต้สมการที่ประมาณได้ e_i มีค่าเป็นลบ

3) สำหรับเส้นสมการที่ผ่านไปในกลุ่มของค่าสังเกต จะเป็นสมการที่ประมาณได้ดีนั้นต้องเป็นสมการที่อาศัยข้อมูลจากทุกค่าสังเกตทุกจุดและต้องสนใจทั้งค่าบวกและค่าลบ หากพิจารณาจากรูปที่ 2.2 จะเห็นว่าส่วนที่เหลือวัดได้จากระยะทางตามแนวตั้ง โดยค่าของ $\hat{\beta}_0, \hat{\beta}_1$ ที่แตกต่างกันจะให้เส้นตรงที่แตกต่างกันและจะให้เซตของส่วนเหลือที่แตกต่างกันไปด้วย ดังนั้นผลรวมกำลังสองของส่วนเหลือ (Residual Sum Square) จึงเป็นฟังก์ชันของ $\hat{\beta}_0, \hat{\beta}_1$ การประมาณค่า $\hat{\beta}_0, \hat{\beta}_1$ ตามแนวคิดของวิธีกำลังสองน้อยสุดสามัญจะต้องเลือก $\hat{\beta}_0, \hat{\beta}_1$ ที่ทำให้ผลรวมกำลังสองของส่วนเหลือมีค่าน้อยที่สุดเท่าที่จะเป็นไปได้ตั้งสมการ

$$RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$RSS = \sum e_i^2 = f(\hat{\beta}_0, \hat{\beta}_1)$$

พิจารณาเงื่อนไขจำเป็น โดยใช้อนุพันธ์บางส่วนกับฟังก์ชัน RSS เทียบกับ $\hat{\beta}_0, \hat{\beta}_1$ จะได้

$$\frac{\partial(\sum e_i^2)}{\partial \widehat{\beta}_0} = -2 \sum (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0$$

$$\frac{\partial(\sum e_i^2)}{\partial \widehat{\beta}_1} = -2 \sum X_i (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0$$

สามารถหาสัมประสิทธิ์ $\widehat{\beta}_1$ ได้จาก

$$\widehat{\beta}_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$\widehat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\widehat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i y_i$$

กำหนดให้ $\sum k_i = \sum \left(\frac{x_i}{\sum x_i^2} \right)$ โดยที่ $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$ และ \bar{X}, \bar{Y} คือค่าเฉลี่ยของข้อมูล

X และ Y ตามลำดับ สามารถหาค่าเฉลี่ย $\widehat{\beta}_1$ โดยใส่ค่าคาดหมายทั้งสองข้างของสมการจะได้

$$E(\widehat{\beta}_1) = \beta_1 + \sum k_i E(\varepsilon_i)$$

เนื่องจากโดยสมมุติ $E(\varepsilon_i) = 0$ ดังนั้น $\widehat{\beta}_1$ จึงเป็นตัวแปรค่าที่ไม่เอนเอียงของ β_1 สำหรับค่าความแปรปรวนของ $\widehat{\beta}_1$ จะหาได้จาก

$$Var(\widehat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

สำหรับสัมประสิทธิ์ $\widehat{\beta}_0$ หาได้จาก

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$$

แทนค่า $\widehat{\beta}_1 = \sum k_i Y_i$ จะได้

$$\widehat{\beta}_0 = \bar{Y} - \bar{X} \sum k_i Y_i$$

$$\widehat{\beta}_0 = \frac{\sum Y_i}{n} - \bar{X} \sum k_i Y_i$$

$$\widehat{\beta}_0 = \sum \left[\frac{1}{n} - \bar{X} k_i \right] Y_i$$

ใส่ค่าคาดหมายทั้งสองข้างของสมการจะได้

$$E(\widehat{\beta}_0) = \sum \left[\frac{1}{n} - \bar{X}k_i \right] E(Y_i)$$

$$E(\widehat{\beta}_0) = \beta_0$$

เนื่องจาก β_0 เป็นค่าเฉลี่ยของ Y เมื่อ $X = 0$ แสดงว่า $\widehat{\beta}_0$ เป็นตัวประมาณที่ค่าไม่มีความเอนเอียงของ β_0 สำหรับความแปรปรวนของ $\widehat{\beta}_0$ จะเท่ากับ

$$Var(\widehat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$$

ในการคำนวณหาค่าความแปรปรวน σ^2 นั้นไม่สามารถหาค่าได้จากค่าของ ε ได้เนื่องจากค่า ε ขึ้นอยู่กับ $\widehat{\beta}_0$ และ $\widehat{\beta}_1$ ซึ่งไม่ทราบค่า แต่สามารถประมาณได้จากการคำนวณส่วนที่เหลือ (Residuals) e_i โดยความแปรปรวนจะมีค่าเท่ากับ

$$\sigma^2 = \frac{\sum e_i^2}{n-2}$$

2.1.8.2 ตัวแบบการถดถอยพหุนาม (Polynomial Regression Models)

ตัวแบบการถดถอยพหุนามมีหลักการพื้นฐานมาจากตัวแบบการถดถอยเชิงเส้น ถ้าพิจารณาตัวแบบพหุนามอันดับสองที่มีตัวแปร 1 ตัว (the second-order polynomial model in one variable) คือ

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

ตัวแบบนี้เรียกว่าตัวแบบอันดับสอง (second-order model) ที่มีตัวแปร 1 ตัว บางครั้งเรียกว่าตัวแบบกำลังสอง (quadratic model) โดยที่ β_0 เป็นค่าเฉลี่ยของ Y เมื่อ $X = 0$ ส่วน β_1 เป็นพารามิเตอร์อิทธิพลเชิงเส้น และ β_2 เป็นพารามิเตอร์อิทธิพลกำลังสอง สามารถเขียนในรูปทั่วไปของตัวแบบพหุนามอันดับที่ k ที่มีตัวแปร 1 ตัวคือ

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k + \varepsilon$$

อันดับของตัวแบบพหุนาม อาจจะหาได้โดยใช้เทคนิคของตัวแบบการถดถอยเชิงเส้นพหุคูณสามารถพิจารณาได้จาก

1. ควรให้อันดับของตัวแบบต่ำที่สุดเท่าที่จะเป็นไปได้ เมื่อฟังก์ชันผลตอบสนองเป็นเส้นโค้ง ควรใช้การแปลงเพื่อให้ตัวแบบเป็นอันดับแรก ถ้าไม่สามารถใช้การแปลงได้ ควรใช้พหุนามอันดับที่สอง
2. วิธีตัวสถิติทดสอบการเลือกไปข้างหน้า (forward selection) คือหาตัวแบบที่มีอันดับเพิ่มขึ้นจนกว่าตัวสถิติทดสอบที่สำหรับเทอมอันดับสูงสุดจะไม่มีนัยสำคัญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. วิธีตัวสถิติทดสอบการกำจัดถอยหลัง (backward elimination) คือหาตัวแบบอันดับสูงสุดและตัดทีละเทอม เริ่มต้นด้วยการตัดอันดับสูงสุดจนกว่าเทอมที่เหลืออยู่อันดับสูงสุดมีตัวสถิติทดสอบที่มีนัยสำคัญ

นอกจากนี้ต้องคำนึงถึงผลของการทำนายกรณีที่อันดับของตัวแบบมีค่าสูงเกินไป เช่น ถ้าทำนายภายนอกพิสัย X ของข้อมูลเริ่มต้น ผลตอบสนองที่ทำนายได้จะลดลง การคำนวณเมทริกซ์จะไม่ถูกต้อง และความคลาดเคลื่อนอาจจะอยู่ในค่าประมาณพารามิเตอร์ ถ้าค่าของ X ถูกจำกัดให้อยู่ในพิสัยแคบโดยมีค่าไม่แตกต่างกันมากนัก จะเกิดความสัมพันธ์เชิงเส้นพหุในแต่ละสมาชิกของเมทริกซ์ X (multicollinearity)

ในงานวิจัยนี้มีการพิจารณาข้อมูลการบำรุงรักษามากกว่า 1 เหตุการณ์ ซึ่งแต่ละเหตุการณ์จะถูกแทนด้วยตัวแปรสุ่ม $X_1, X_2, X_3, \dots, X_n$ ดังนั้นตัวแบบพหุนามที่ใช้ในการคำนวณจะมีตัวแปร 2 ตัว หรือมากกว่า 2 ตัว (Polynomial Models in Two or More Variables) เช่น ตัวแบบพหุนามอันดับสองที่มีตัวแปร 2 ตัว คือ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

2.1.8.3 การแบ่งกลุ่มข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล (Classification)

(นิเวศ, 2010) เนื่องจากการทำเหมืองข้อมูลเป็นเทคนิคในการค้นคว้าความรู้จากข้อมูลขนาดใหญ่ การทำเหมืองข้อมูลจึงเป็นการรวมองค์ความรู้หลายส่วน เช่น เทคโนโลยีฐานข้อมูล วิทยาศาสตร์สารสนเทศ สถิติ และระบบการเรียนรู้ เป็นต้น ในเรื่องของการทำงานและการจัดกลุ่มของข้อมูลนั้นเป็นการทำเหมืองข้อมูลรูปแบบหนึ่ง จัดเป็นกระบวนการที่ใช้ในการหารูปแบบของชุดข้อมูลที่มีความใกล้เคียงกันหรือเหมือนกันมากที่สุด เพื่อใช้ในการทำนายชุดข้อมูลว่าอยู่ในประเภทใดของชุดข้อมูลที่ได้ทำการแบ่งไว้แล้ว ซึ่งชุดข้อมูลที่แบ่งไว้เกิดจากการเรียนรู้จากชุดข้อมูลฝึกหัด

(Roy et al., 2021) ได้นำเทคนิคการเรียนรู้ของเครื่องมาใช้ในการจัดกลุ่มข้อมูลความผิดปกติ 2 เหตุการณ์ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน แต่ในงานวิจัยนี้พิจารณาความผิดปกติมากกว่า 2 เหตุการณ์ จึงเลือกใช้วิธีต้นไม้ตัดสินใจแทนเพื่อลดความซับซ้อนของการหาเส้นแบ่งในวิธีซัพพอร์ตเวกเตอร์แมชชีน

1. วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

วิธีการนี้ใช้เพื่อหาระนาบการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยใช้สมการเส้นตรงเพื่อแบ่งเขตข้อมูล 2 กลุ่ม ออกจากกัน โดยมีวัตถุประสงค์ที่จะพยายามที่จะทำการลดความผิดพลาดจากการทำนายพร้อมกับเพิ่มระยะแยกแยะให้มากที่สุด ซึ่งต่างจากเทคนิคโดยทั่วไป เช่น โครงข่ายประสาทเทียมที่มุ่งเน้นการลดความผิดพลาดจากการทำนายให้ต่ำที่สุดเพียงอย่างเดียว โดยใช้ฟังก์ชันแปลงข้อมูลป้อนเข้าให้อยู่ในปริภูมิคุณลักษณะ (Feature Space) และสร้างฟังก์ชันเคอร์

เนล (Kernel Function) สำหรับวัดความคล้ายของข้อมูลในปริภูมิคุณลักษณะ เหมาะสำหรับข้อมูลที่
มีมิติของข้อมูลปริมาณมากกำหนดให้ (Pingwu and Dewang, 2009)

$$D = \{(X_i, y_i); i = 1, 2, 3, \dots, n\}$$

$$\text{โดยที่ } X_i = (x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_m}) \in R^m \text{ และ } y_i \in \{-1, 1\}$$

ในการแบ่งข้อมูลเชิงเส้น มิติข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยระนาบตัดสินใจ ซึ่ง
คำนวณได้ดังสมการ

$$w^T X + b \geq y - \xi_i; y = 1$$

$$w^T X + b \geq y + \xi_i; y = -1$$

กำหนดให้ $\xi_i \geq 0$ จะสามารถหาเส้นแบ่งที่มีข้อผิดพลาดน้อยที่สุดได้จาก w คือ ค่าน้ำหนักและ
 b คือค่า bias อย่างไรก็ตาม ซัพพอร์ตเวกเตอร์แมชชีนมีฟังก์ชันเคอร์เนล ที่ผู้วิจัยสามารถประยุกต์ใช้
ในการแก้ปัญหาให้เหมาะสมกับลักษณะข้อมูลหลายรูปแบบ เช่น ฟังก์ชันพหุนาม ฟังก์ชันเรเดียล
เบสิส ฟังก์ชันซิกมอยด์ เป็นต้น

2. วิธีต้นไม้ตัดสินใจ (Decision Tree)

วิธีการนี้ใช้เพื่อหาสมบัติและค่าที่เหมาะสมในการแบ่งกลุ่มของข้อมูลตามลักษณะต้นไม้ จะ
ประกอบด้วยโหนดแทนคุณลักษณะ และโหนดล่างสุดแทนหมวดหมู่ การสร้างกิ่งสาขาจะพิจารณา
จากค่าความจริงของคุณลักษณะ โดยการสร้างต้นไม้ตัดสินใจ C4.5 จะคำนวณจากค่าเกนสารสนเทศ
(Information Gain) เพื่อเลือกคุณลักษณะที่จะใช้เป็นรากหรือโหนด ถ้าให้ชุดข้อมูล M ประกอบด้วย
ค่าที่เป็นไปได้ คือ $m_1, m_2, m_3, \dots, m_n$ และให้ความน่าจะเป็นที่จะเกิดค่า m_i มีค่าเท่ากับ $P(m_i)$ จะได้
ว่าค่าเกนสารสนเทศของ M เขียนแทนด้วย $I(M)$ คำนวณได้ดังนี้ (นิเวศ, 2010)

$$I(M) = \sum_{i=1}^n -P(m_i) \log_2 P(m_i)$$

ถ้ากำหนดข้อมูลฝึก T คุณลักษณะที่เป็นโหนด x และมีค่าทั้งหมดที่เป็นไปได้ n ค่า โหนด
ปัจจุบันจะแบ่งตัวอย่าง T ออกตามกิ่งเป็น $t_1, t_2, t_3, \dots, t_n$ ตามค่าที่เป็นไปได้ของ x ดังนั้นจึงสามารถ
คำนวณค่าเกนสารสนเทศหลังจากการแบ่งตามคุณลักษณะและค่ามาตรฐานเกนของคุณลักษณะ x ได้
ดังสมการ

$$I_x(T) = \sum_{i=1}^n \frac{|t_i|}{|T|} I(t_i)$$

$$\text{Gain} = I(T) - I_x(T)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นคำนวณค่าสารสนเทศของการแบ่งแยกคุณลักษณะแต่ละตัว ถ้าให้ T คือ ชุดของตัวอย่าง เมื่อแบ่งตัวอย่างนี้ตามคุณลักษณะ x จะได้ชุดของตัวอย่างย่อยในแต่ละกิ่ง คือ $t_1, t_2, t_3, \dots, t_n$ จำนวน n ชุด ตามค่าที่เป็นไปได้ในสมบัติ x เมื่อคำนวณค่าสารสนเทศของการแบ่งแยกได้ดังสมการ

$$\text{Split Information} = \sum_{i=1}^n \frac{|t_i|}{|T|} \log_2 \frac{|t_i|}{|T|}$$

คำนวณค่ามาตรฐานอัตราส่วนเกน (Gain ratio) ได้จาก

$$\text{Gain Ratio} = \text{Gain} - \text{Split Information}$$

ท้ายสุดจึงเลือกค่าอัตราส่วนเกนสูงสุดเป็นคุณลักษณะเริ่มต้น และเลือกสมบัติถัดไปตามค่า Gain ratio น้อยลงตามลำดับ

2.1.9 การพัฒนาตัวแบบในข้อมูลการใช้งานเครื่องจักร

ในงานวิจัยนี้นำแนวคิดของแผนภูมิควบคุมข้อมูลผันแปร (Variable Control Chart) มาใช้ในการหาค่าความผิดปกติของข้อมูล (Liu and Tatera, 2006) แผนภูมิควบคุมข้อมูลผันแปร คือ แผนภูมิควบคุมข้อมูลวัดที่มีลักษณะต่อเนื่องและมีลักษณะผันแปร (Variable) เช่น เส้นผ่านศูนย์กลาง ความแข็ง แรงต้าน แรงดึง ค่าแรงดันไฟฟ้า ค่าใช้จ่าย เป็นต้น จึงเรียกแผนภูมิควบคุมข้อมูลเหล่านี้ว่า แผนภูมิควบคุมข้อมูลผันแปร โดยทั่วไปนิยมใช้กัน 3 ประเภทคือ แผนภูมิควบคุมค่าเฉลี่ยและพิสัย ($\bar{X} - R$) แผนภูมิควบคุมค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ($\bar{X} - S$) และแผนภูมิควบคุมค่าเฉลี่ยและพิสัยเคลื่อนที่ ($\bar{X} - MR$) แผนภูมิทั้งสามประเภทนี้มีวัตถุประสงค์หลักเพื่อการควบคุมแนวโน้มเข้าสู่ศูนย์กลางหรือค่าเฉลี่ยและการกระจายของลักษณะคุณภาพ โดยแผนภูมิ \bar{X} ใช้เพื่อควบคุมค่าเฉลี่ยของลักษณะคุณภาพ ส่วนแผนภูมิ R แผนภูมิ S และแผนภูมิ MR ใช้เพื่อควบคุมการกระจายของลักษณะคุณภาพ

1. แผนภูมิควบคุมค่าเฉลี่ยและพิสัย ($\bar{X} - R$) เป็นแผนภูมิประเภทข้อมูลผันแปรที่มีความเหมาะสมกับกระบวนการที่มีสมบัติการเกิดซ้ำค่อนข้างต่ำและขนาดกลุ่มตัวอย่างมีขนาดเล็ก หากข้อมูลมีการกระจายน้อยแสดงว่าค่าสมบัติของผลิตภัณฑ์มีค่าใกล้เคียงกัน ในการเก็บข้อมูลจะพิจารณาเป็นกลุ่มย่อยและมีการคำนวณขีดจำกัดควบคุมดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \bar{\bar{X}} + A_2 \bar{R}$$

$$\text{เส้นกึ่งกลาง} = \bar{\bar{X}}$$

$$\text{ขีดจำกัดควบคุมล่าง} = \bar{\bar{X}} - A_2 \bar{R}$$

โดยที่ $\bar{\bar{X}}$ คือค่าเฉลี่ยของค่าเฉลี่ยของข้อมูลในกลุ่มย่อย \bar{R} คือค่าเฉลี่ยของพิสัยของข้อมูลในกลุ่มย่อย และ A_2 เป็นค่าคงที่ขึ้นอยู่กับขนาดตัวอย่าง n และพิสัยในแต่ละกลุ่มย่อย

2. แผนภูมิควบคุมค่าเฉลี่ยและพิสัยเคลื่อนที่ ($\bar{X} - MR$) เป็นแผนภูมิประเภทข้อมูลผันแปรที่มีความเหมาะสมกับกระบวนการที่มีสมบัติการเกิดซ้ำสูง จนไม่มีความสนใจต้องควบคุมในการเก็บข้อมูลจึงไม่จำเป็นต้องพิจารณาเป็นกลุ่มย่อย มีการคำนวณขีดจำกัดควบคุมดังนี้

$$\text{ขีดจำกัดควบคุมบน} = D_4 \overline{MR}$$

$$\text{เส้นกึ่งกลาง} = \overline{MR}$$

$$\text{ขีดจำกัดควบคุมล่าง} = D_3 \overline{MR}$$

โดยที่ \overline{MR} คือค่าเฉลี่ยของพิสัยเคลื่อนที่ของข้อมูล

$$\overline{MR} = \frac{MR_1 + MR_2 + \dots + MR_{k-1}}{k-1}$$

และ D_3, D_4 เป็นค่าคงที่ขึ้นอยู่กับขนาดตัวอย่าง n ในแต่ละกลุ่มย่อย

3. แผนภูมิควบคุมค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ($\bar{X} - S$) เป็นแผนภูมิประเภทข้อมูลผันแปรที่มีความเหมาะสมกับกระบวนการที่มีสมบัติการเกิดซ้ำต่ำ (Repeatability) แต่แตกต่างกับแผนภูมิควบคุมค่าเฉลี่ยและพิสัยที่ขนาดกลุ่มย่อยของข้อมูลต้องมีมากกว่า 9 ตัว มีการคำนวณขีดจำกัดควบคุมดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \bar{\bar{X}} + A_3 \bar{S}$$

$$\text{เส้นกึ่งกลาง} = \bar{\bar{X}}$$

$$\text{ขีดจำกัดควบคุมล่าง} = \bar{\bar{X}} - A_3 \bar{S}$$

โดยที่ $\bar{\bar{X}}$ คือค่าเฉลี่ยของค่าเฉลี่ยของข้อมูลในกลุ่มย่อย \bar{S} คือค่าเฉลี่ยของส่วนเบี่ยงเบนมาตรฐานของข้อมูลในกลุ่มย่อย

$$\bar{S} = \frac{S_1 + S_2 + \dots + S_k}{k}$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

เมื่อ k คือจำนวนกลุ่มย่อย X_i คือค่าที่วัดได้ของแต่ละกลุ่มย่อย n คือขนาดตัวอย่างในแต่ละกลุ่มย่อย และ A_3 เป็นค่าคงที่ขึ้นอยู่กับขนาดตัวอย่าง n ในแต่ละกลุ่มย่อย สังเกตได้ว่าแผนภูมิควบคุมทั้ง 3 แผนภูมิที่กล่าวไปจะพิจารณาข้อมูลผันแปรเพียงตัวเดียว ในงานวิจัยจำนวนมากจึงได้พัฒนาแผนภูมิควบคุมสำหรับข้อมูลผันแปรหลายตัวแปรขึ้น เช่น

2.1.9.1 (Hayati, 2017) ได้เสนอวิธีกำหนดแผนภูมิควบคุมการถดถอย (Regression control chart) ซึ่งพัฒนามาจากแผนภูมิควบคุมทั่วไปและการถดถอยเชิงเส้นสำหรับข้อมูล 2 ตัวแปรที่มีความสัมพันธ์กัน วิธีกำหนดขีดจำกัดควบคุมมีดังนี้

1. พิจารณาตัวแปรต้นและตัวแปรตามต้องมีความสัมพันธ์เชิงเส้น $m(X, Y)$
2. พิจารณาระดับความสัมพันธ์ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (R) เมื่อพิจารณา $t = R\sqrt{n}$ โดยที่ n คือขนาดของชุดข้อมูล หาก $t > 3$ จะแสดงถึงความสัมพันธ์ที่มีนัยสำคัญ
3. ประมาณค่าสมการถดถอยเชิงเส้น
4. หาค่าความคลาดเคลื่อนมาตรฐาน (Standard error) จากส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามดังสมการ $S_{error} = S.D._Y \sqrt{1-r^2}$

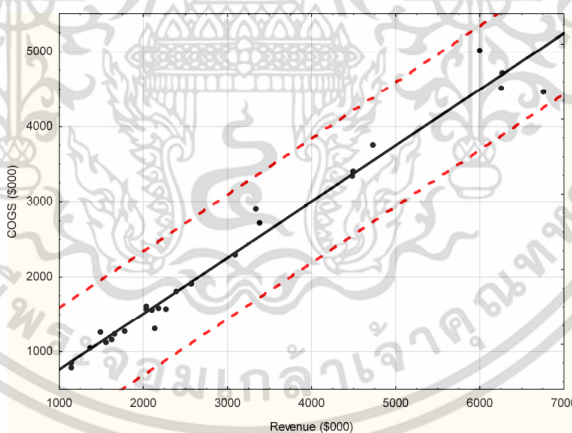
5. กำหนดค่าขีดจำกัดควบคุมดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \hat{Y}_i + 3 S_{error}$$

$$\text{เส้นกึ่งกลาง} = \hat{Y}_i \text{ หรือ ค่าทำนายจากสมการถดถอยเชิงเส้น} = \beta_0 + \beta_1 X$$

$$\text{ขีดจำกัดควบคุมล่าง} = \hat{Y}_i - 3 S_{error}$$

ทางผู้วิจัยได้นำหลักการดังกล่าวไปใช้ประโยชน์ทางด้านธุรกิจหลายด้าน เช่น การควบคุมต้นทุนการผลิตเทียบกับรายได้ ดังรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างแผนภูมิควบคุมการถดถอย

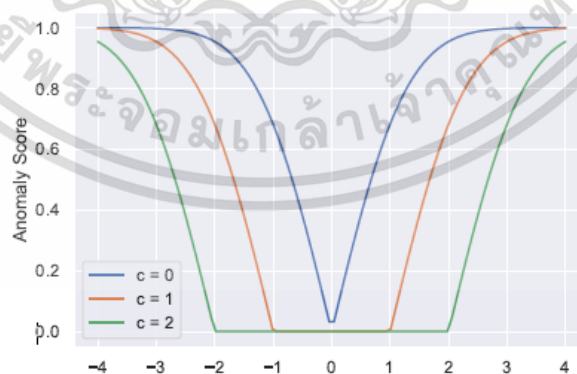
จากรูปที่ 2.3 แสดงให้เห็นถึงขอบเขตการควบคุมต้นทุนการผลิตเทียบกับรายได้ เมื่อเส้นกึ่งกลางคือต้นทุนการผลิตที่ได้จากการประมาณค่าของรายได้ หากจุดข้อมูลที่เกิดขึ้นใหม่อยู่ในกรอบขีดจำกัดตามเส้นประสีแดงหมายถึงยังไม่พบความผิดปกติของต้นทุนการผลิตเทียบกับรายได้

2.1.9.2 (Cetin and Tasgin, 2021) ได้ศึกษาวิธีการตรวจจับความผิดปกติ พบว่าความผิดปกติเป็นเหตุการณ์ที่เกิดขึ้นได้ยาก อย่างไรก็ตามการเกิดความผิดปกติเพียงกรณีเดียวอาจมีค่าใช้จ่ายสูงในหลายมุมมอง เช่น การสูญเสียทางการเงิน การละเมิดความปลอดภัย เป็นต้น ปัญหาหนึ่งสำหรับวิธีการตรวจหาสิ่งผิดปกติคือการไม่มีชุดข้อมูลที่ถูกระบุความผิดปกติที่ครบถ้วน กล่าวคือข้อมูลส่วนใหญ่ไม่มีการระบุความผิดปกติกำกับไว้ และพฤติกรรมของความผิดปกติที่เปลี่ยนแปลงอย่างต่อเนื่อง ทำให้ข้อมูลไม่เป็นปัจจุบันสำหรับพัฒนาต้นแบบ เทคนิคหนึ่งในการตรวจจับความผิดปกติทางสถิติที่ง่ายและมีประสิทธิภาพคือวิธี k -sigma วิธีพิจารณามีดังนี้

- 1) เตรียมชุดของจุดข้อมูล X จำนวน i อันดับ
- 2) คำนวณค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ ของจุดข้อมูล X ทั้งหมด
- 3) เปรียบเทียบถ้าจุดข้อมูล x_i ที่เป็นสมาชิกเซต X หากตกอยู่นอก k เท่าของส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย นั่นคือ $|x_i - \mu| > k\sigma$ จะถูกจัดว่าเป็นจุดข้อมูลผิดปกติ มิฉะนั้น x_i จะจัดเป็นจุดข้อมูลปกติ
- 4) คำนวณคะแนนความผิดปกติ ซึ่งจะวัดระดับของค่าผิดปกติหรือระดับความผิดปกติสำหรับจุดข้อมูลแต่ละจุดข้อมูลในชุดข้อมูลที่กำหนด หากชุดข้อมูลมีตัวแปรสุ่มเพียงตัวเดียวและมีการแจกแจงปรกติจะสามารถหาคะแนนความผิดปกติได้จาก

$$\text{score} = \begin{cases} s(k_i - c), k_i = \frac{|X - \bar{X}|}{\sigma} \geq c \\ 0, \text{otherwise} \end{cases}$$

โดยที่ s คือความน่าจะเป็นสะสมของการแจกแจงปรกติ และ c เป็นค่าคงที่ของช่วงการเบี่ยงเบนจากค่าเฉลี่ยที่ยอมรับได้ ซึ่งสามารถแสดงค่าคะแนนได้ตามรูปที่ 2.4



รูปที่ 2.4 การแจกแจงความน่าจะเป็นสะสมของการแจกแจงปรกติเทียบกับช่วงการเบี่ยงเบนจากค่าเฉลี่ยที่ยอมรับได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.4 เมื่อกำหนดค่าคงที่ $c = 2$ จะเห็นว่าค่าคะแนนความผิดปกติจะมีค่าเป็น 0 เมื่อจุดของข้อมูลอยู่ในช่วง $(-2\sigma, 2\sigma)$ และจะมีค่าคะแนนความผิดปกติตามความน่าจะเป็นสะสมของการแจกแจงปกติเมื่อจุดของข้อมูลอยู่นอกช่วง $(-2\sigma, 2\sigma)$

นอกจากนี้กรณีที่ชุดข้อมูลมีตัวแปรสุ่มหลายตัว ทางผู้วิจัยได้นำวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร (Multi-variate k -sigma) มาหาระยะทางจากค่าเฉลี่ยรวมของ 2 ตัวแปรสามารถคำนวณได้จากระยะทางมาฮาลาโนบิส

$$k_i = \sqrt{(\bar{X}_i - \bar{\pi})^T \Sigma^{-1} (\bar{X}_i - \bar{\pi})}$$

โดยที่ \bar{X}_i คือเวกเตอร์ของจุดข้อมูลในตัวแปรสุ่ม $\bar{\pi}$ คือเวกเตอร์ค่าเฉลี่ยของจุดข้อมูลในตัวแปรสุ่ม และ Σ คือเมทริกซ์ความแปรปรวนร่วมของจุดข้อมูล

อีกทั้งทางผู้วิจัยได้เปรียบเทียบกับวิธีป่าไม้โดดเดี่ยว (Isolation forest) ซึ่งเป็นวิธีการเรียนรู้ของเครื่องแบบไม่มีผู้สอน มีหลักการบนพื้นฐานของจัดกลุ่มด้วยต้นไม้ตัดสินใจหลายต้นเพื่อลดค่าความแปรปรวนของผลทำนาย โดยต้นไม้แต่ละต้นมีเป้าหมายคือลดค่าความสัมพันธ์ระหว่างข้อมูลจนได้กลุ่มที่มีสมาชิกที่น้อยที่สุด แล้วพิจารณาจำนวนโหนดตัดสินใจ หากค่าสังเกตมีจำนวนโหนดตัดสินใจเฉลี่ยน้อยจะระบุว่าข้อมูลนั้นเป็นค่าที่ผิดปกติจากชุดข้อมูลฝึก วิธีนี้สามารถหาความผิดปกติได้ทั้งตัวแปรสุ่ม 1 ตัวแปร และหลายตัวแปร วิธีป่าไม้โดดเดี่ยวแตกต่างจากวิธีตามระยะทางและวิธีตามความหนาแน่นดังนี้

1. ลักษณะการแบ่งข้อมูลเป็นส่วนย่อย ให้ต้นไม้ตัดสินใจจำนวนมากช่วยสร้างแบบจำลอง ทำให้สามารถใช้ประโยชน์จากการสุ่มข้อมูลส่วนย่อยในการแก้ปัญหาการทำนายชุดข้อมูลขนาดใหญ่ เช่น ทำนายข้อมูลปกติเป็นผิดปกติ (Swamping) และการรวมข้อมูลผิดปกติเป็นข้อมูลปกติในขั้นตอนการสร้างแบบจำลอง (Masking)
2. ในการคำนวณจะไม่ใช้การหาระยะทางหรือความหนาแน่นในการตรวจจับข้อมูลผิดปกติ ทำให้สามารถลดเวลาและหน่วยความจำ
3. เมื่อข้อมูลมีขนาดใหญ่ขึ้น ความซับซ้อนในการคำนวณจะไม่เพิ่มขึ้นตามจำนวนข้อมูลเป็นเชิงเส้น
4. มีความสามารถในการสร้างต้นแบบที่ได้รับผลกระทบจากข้อมูลขนาดใหญ่และมีคุณลักษณะที่ไม่เกี่ยวข้องจำนวนมาก

2.1.10 การพิจารณาตัวแบบที่เหมาะสม (Davide et al., 2021)

2.1.10.1 ความเหมาะสมของพารามิเตอร์ที่ใช้ในตัวแบบ

1. ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient, R) เป็นค่าที่แสดง ถึงระดับความสัมพันธ์ระหว่างกลุ่มของตัวแปรอิสระทั้งหมดในสมการกับตัวแปรตาม ค่าเข้าใกล้ 1 แสดงว่ากลุ่มของตัวแปรอิสระมีความสัมพันธ์กับตัวแปรตามมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ค่าสัมประสิทธิ์การกำหนดพหุคูณ (Coefficient of Multiple Determination, R Square) เป็นค่าที่แสดงความสัมพันธ์ตัวแปรอิสระทั้งหมดในสมการที่มีต่อตัวแปรตามทั้งหมดที่อยู่ในสมการสามารถอธิบายการผันแปรตัวแปรตามได้มากน้อยเพียงใด โดยอธิบายว่าตัวแปรอิสระทั้งหมดมีอิทธิพลต่อตัวแปรตามอยู่ร้อยละค่าสัมประสิทธิ์การกำหนดพหุคูณ ส่วนที่เหลือจะเป็นอิทธิพลจากตัวแปรอื่นที่ไม่ได้นำมาวิเคราะห์หรือตัวแปรอิสระทั้งหมดสามารถทำนายตัวแปรตามได้ร้อยละค่าสัมประสิทธิ์การกำหนดพหุคูณ

3. ค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้ว (Adjusted R Square) ควรพิจารณาใช้ค่านีแทนค่าสัมประสิทธิ์การกำหนดพหุคูณเมื่อข้อมูลที่นำมาวิเคราะห์สมการถดถอยมีค่าน้อยกว่า 30 ตัวอย่าง หรือในกรณีที่ค่าสัมประสิทธิ์การกำหนดพหุคูณจะมีค่าสูงมากกว่าค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้วมาก จึงควรใช้ค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้วแทน

4. ความคลาดเคลื่อนมาตรฐานของค่าประมาณ (Standard Error of Estimate) คือ ค่าที่แสดงระดับของความคลาดเคลื่อนที่เกิดจากการใช้ตัวแปรอิสระทั้งหมดมาทำนายตัวแปรตาม การพิจารณาค่าความคลาดเคลื่อนนี้จะพิจารณาจากหน่วยที่ใช้ด้วยจึงจะสามารถบอกได้ว่าความคลาดเคลื่อนนี้สูงหรือต่ำ

2.1.10.2 การประเมินค่าความผิดพลาดจากการทำนาย

การประเมินประสิทธิภาพของการทำนายค่าข้อมูลที่อยู่ในรูปตัวเลข จะใช้ค่าความผิดพลาดที่เกิดขึ้นของการทำนายชุดข้อมูลทดสอบโดยเปรียบเทียบค่าผลลัพธ์ที่แท้จริงกับค่าผลลัพธ์จากการทำนาย ค่าความผิดพลาดยังมีค่าน้อยหมายถึงตัวแบบมีความเที่ยง กำหนดให้ Y คือค่าข้อมูลที่แท้จริง \hat{Y} คือ ค่าผลลัพธ์ที่ได้จากการทำนาย n คือจำนวนข้อมูลทั้งหมด วิธีการคำนวณมีดังนี้

1. ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE)

$$MSE = \frac{\sum (Y_i - \hat{Y}_i)^2}{n}$$

2. รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error, RMSE)

$$RMSE = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n}}$$

3. ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error, MAE)

$$MAE = \frac{\sum |Y_i - \hat{Y}_i|}{n}$$

4. ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error, MAPE)

$$MAPE = \frac{100}{n} \left[\sum \frac{|Y_i - \hat{Y}_i|}{Y_i} \right]$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 งานวิจัยที่เกี่ยวข้อง

Wang and Yin (2019) ศึกษาเกี่ยวกับการผสมผสานการแจกแจงไวบูลกับการวิเคราะห์อนุกรมเวลา ทำนายอัตราความผิดปกติของสถานีไฟฟ้าย่อยเทียบกับเวลา ทำการเปรียบเทียบการวิเคราะห์อนุกรมเวลาร่วมกับการแจกแจงไวบูลกับการแจกแจงไวบูลเพียงอย่างเดียว ด้วยค่าส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ย (mean absolute deviation) ผลการศึกษาพบว่า การวิเคราะห์อนุกรมเวลาร่วมกับการแจกแจงไวบูลนั้นมีค่าส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ยต่ำกว่าการแจกแจงไวบูลเพียงอย่างเดียว โดยเฉลี่ย 17% เนื่องจากในงานวิจัยมีการวิเคราะห์อัตราความผิดปกติด้วยอนุกรมเวลาด้วย ทำให้ต้องมีการประมาณค่าอัตราความผิดปกติในแต่ละช่วงเวลา ซึ่งทางผู้วิจัยได้ใช้วิธีการประมาณด้วยวิธีสไปรน์แบบลูกบาศก์จากนั้นนำมาตรวจสอบค่าอัตราความผิดปกติเทียบกับอัตราความผิดปกติรูปอ่าง เพื่อหาจุดแบ่งคาบระหว่าง อัตราความผิดปกติคงที่ตามช่วงเวลามีลักษณะสุมและอัตราความผิดปกติที่ขึ้นอยู่กับช่วงเวลา ด้วยวิธีผลรวมกำลังสองของส่วนเหลือที่น้อยที่สุด ในส่วนของความผิดปกติคงที่ตามช่วงเวลาได้นำการวิเคราะห์อนุกรมเวลาไปใช้ในการทำนายอัตราความผิดปกติ ทางผู้วิจัยได้มีการตรวจสอบองค์ประกอบของอนุกรมเวลา และทดสอบความเสถียรด้วยสถิติทดสอบ ADF พบว่าข้อมูลมีรูปแบบคงที่ ไม่มีแนวโน้มการเปลี่ยนแปลง จึงพิจารณาประมาณต้นแบบด้วยวิธี ARMA ด้วยวิธีค่าเฉลี่ยเป็นศูนย์เพื่อตัดค่าคงที่ออกจากข้อมูลก่อนนำไปสร้างต้นแบบ ทำการประมาณค่าตัวแปรของ ARMA ด้วยวิธีอาเคเค Akaike Information Criterion (AIC) พิจารณาตัวที่มีค่า AIC ต่ำสุดก่อน ในส่วนอัตราความผิดปกติที่ขึ้นอยู่กับช่วงเวลา ได้นำการวิเคราะห์การแจกแจงไวบูลมาใช้ในการทำนายอัตราความผิดปกติ

Roy et al. (2021) ศึกษาการแจกแจงความผิดปกติของเครื่องจักรกังหัน โดยเปรียบเทียบการแจกแจงแบบเดียว (การแจกแจงล็อกปรกติ การแจกแจงแกมมา และการแจกแจงไวบูล) การแจกแจงประกอบ และการแจกแจงผสม ด้วยสถิติทดสอบคอลโมโกรอฟ-สมิรโนฟ พบว่าการแจกแจงเชิงประกอบได้ค่า K-S น้อยที่สุด ในงานวิจัยนี้เปรียบเทียบการแจกแจงแบบเดียวกับข้อมูลความผิดปกติของเครื่องจักร ซึ่งแบ่งได้เป็น 2 รูปแบบคือความผิดปกติของเครื่องจักรที่ทำการบำรุงรักษาใหญ่และความผิดปกติของเครื่องจักรที่ยังไม่ทำการบำรุงรักษาใหญ่ ทำการเปรียบเทียบการทดสอบภาวะสารูปตีวิธีสถิติทดสอบคอลโมโกรอฟ-สมิรโนฟ ได้ผลว่าการแจกแจงไวบูลให้ค่า K-S น้อยที่สุดในรูปแบบความผิดปกติของเครื่องจักรที่ทำการบำรุงรักษาใหญ่และการแจกแจงแกมมาให้ค่า K-S น้อยที่สุดในความผิดปกติของเครื่องจักรที่ยังไม่ทำการบำรุงรักษาใหญ่ ทางผู้วิจัยจึงเสนอรวมทั้ง 2 รูปแบบความผิดปกติ และเสนอวิธีการแจกแจงเชิงประกอบจากการรวมการแจกแจงไวบูลและการแจกแจงแกมมาเข้าด้วยกัน และมีจุดแบ่งการแจกแจงคือ t_1 ซึ่งจุดแบ่งดังกล่าวทางผู้วิจัยได้นำเทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) มาใช้ในการแบ่งกลุ่มของข้อมูล นอกจากนี้ทางผู้วิจัยได้เสนอการแจกแจงผสม (Mixed Distribution) ซึ่งพัฒนามาจากการแจกแจงประกอบ เพิ่มเติมในส่วนของคุณสมบัติการถดถอยของแต่ละการแจกแจง (W_1, W_2) ในแต่ละช่วงของเวลาที่จุดแบ่งเดียวกัน (t_1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Cetin and Tasgin 2021 ได้กล่าวถึงการวิเคราะห์ความผิดปกติด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานและวิเคราะห์ความผิดปกติหลายตัวแปรด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร ประกอบกับวิธีการเรียนรู้ของเครื่องอีก 2 วิธี คือแบบจำลองป่าไม้โคตเต็ย และ วิธีการจัดกลุ่มเฉลี่ย k กลุ่ม บนข้อมูลสังเคราะห์ ทำการเปรียบเทียบทั้ง 4 วิธีด้วยค่าเรียกคืน ความเที่ยง และ คะแนน F1 พบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร ได้ค่าสูงที่สุดทั้งค่าเรียกคืน ความเที่ยง และ คะแนน F1 ผู้วิจัยได้ทดลองนำข้อมูลสังเคราะห์ 2 ตัวแปรจำนวน 7 กลุ่ม รวม 3,000 ตัวอย่าง ด้วยฟังก์ชัน `make_blobs` ใน `sklearn` package และเพิ่มข้อมูลสุ่มแทนข้อมูลที่ผิดปกติจำนวน 100 ตัวอย่าง ด้วยการประมาณค่าอนติคาร์โลเพื่อจำลองข้อมูลเพิ่มเติมด้วยเงื่อนไขค่าเฉลี่ยและความแปรปรวนร่วมเดิม และยังช่วยลดความซับซ้อนในการประมวลผลทำให้เวลาในการคำนวณสั้นลง

Yenel and Jimenez 2020 ศึกษาเกี่ยวกับการหาแบบจำลองความผิดปกติเพื่อใช้กำหนดระยะเวลาการรับประกันของอุตสาหกรรมยานยนต์ จะพิจารณาระยะทางที่ยานยนต์ใช้งาน (mileage) ร่วมกับการวิเคราะห์ความผิดปกติของอุปกรณ์ เทียบกับการวิเคราะห์ความผิดปกติของอุปกรณ์เพียงอย่างเดียว ด้วยวิธีส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ย (mean absolute deviation) ผลการศึกษาพบว่าวิธีการแจกแจงความน่าจะเป็นร่วม (Joint Distribution) ระหว่างระยะทางที่ยานยนต์ใช้งานกับการวิเคราะห์ความผิดปกติของอุปกรณ์ต่ำกว่าการแจกแจงความผิดปกติเพียงอย่างเดียวที่ 0.949 หน่วย และค่าเอนเอียง (bias) ที่ -0.32 หน่วย โดยทั่วไปการวิเคราะห์ความผิดปกติของอุปกรณ์จะดำเนินการในช่วงเวลาหรือขอบเขตของอายุการใช้งาน ในบางกรณีผลกระทบของตัวแปรอื่นที่มีต่อความน่าเชื่อถือของอุปกรณ์จะถูกนำมาพิจารณาด้วย แม้ว่าจะเป็นเรื่องง่ายในการวิเคราะห์ฟังก์ชันความหนาแน่นความน่าจะเป็นแบบไม่มีเงื่อนไขอย่างอิสระ แต่ปัญหาของความเป็นอิสระทางสถิติระหว่างตัวแปรสุ่มของการใช้งานและความผิดปกติของยานยนต์ซึ่งมีข้อมูลเวลาร่วมกันนั้น ฟังก์ชันความน่าจะเป็นร่วมจึงไม่สามารถเป็นผลคูณของฟังก์ชันความหนาแน่นของความน่าจะเป็นส่วนเพิ่มทั้งสองได้ เนื่องจากตัวแปรสุ่มทั้งสองมีความสัมพันธ์ในอัตราส่วนของเวลาและมีการแจกแจงล็อกนอร์มอล ดังนั้นสามารถหาความน่าจะเป็นร่วมได้จากความน่าจะเป็นแบบมีเงื่อนไขของเวลาระหว่างการแจกแจงไวบูลและการแจกแจงล็อกปรกติ

บทที่ 3

วิธีการดำเนินงานวิจัย

ข้อมูลในงานวิจัยนี้ประกอบด้วยข้อมูลการบำรุงรักษาเครื่องจักรมีการแจกแจงไวบูล (Weibull distribution) และข้อมูลการใช้งานเครื่องจักรมีการแจกแจงปกติ (Normal distribution)

3.1 ข้อมูลการบำรุงรักษาเครื่องจักร (Maintenance Data)

ในงานวิจัยนี้ได้พัฒนาตัวแบบจากข้อมูลการบำรุงรักษาเครื่องจักร จากประวัติการซ่อมและประวัติความผิดปกติของแต่ละเครื่องจักร เช่น อาการความผิดปกติ วันเวลาที่พบความผิดปกติ และข้อมูลรหัสอุปกรณ์ เป็นต้น ตามขั้นตอนต่อไปนี้

3.1.1 กำหนดขนาดตัวอย่างข้อมูลการบำรุงรักษาของเครื่องจักรทั้งหมดในสถานีเพิ่มความดันก๊าซธรรมชาติ จำนวน 119 ตัวอย่าง ตั้งแต่วันที่ 1 พ.ค. 2563 ถึง 20 เม.ย. 2565

3.1.2 เตรียมข้อมูลและตรวจสอบการแจกแจงไวบูล

1) นำข้อมูลประวัติการบำรุงรักษามาคำนวณอายุการใช้งานเครื่องจักร (Time to fail)

TTF = วันที่เครื่องจักรผิดปกติ - วันที่เครื่องจักรพร้อมใช้งานครั้งล่าสุด

2) เรียงลำดับ TTF จากน้อยไปมากและกำหนดค่าอันดับที่ของมัธยฐาน ตามการประมาณค่า

ของปีนาร์ต ซึ่งจะแสดงค่าความน่าจะเป็นสะสมของความผิดปกติ $F_i(t) = \frac{i - 0.3}{n + 0.4}$

เมื่อ i คืออันดับที่ของ TTF ที่เรียงลำดับจากน้อยไปมาก และ n คือขนาดตัวอย่าง

3) ประมาณค่าพารามิเตอร์การแจกแจงไวบูล ประกอบด้วยตัวแปรบ่งขนาดหรือ β และตัวแปรรูปร่างหรือ α ด้วยวิธี Curve fitting ด้วยสมการความน่าจะเป็นสะสมของการแจกแจงไวบูล

$$F(t; \alpha, \beta) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha} \quad (3.1)$$

$$\ln\left(\ln\left(\frac{1}{1-F(t)}\right)\right) = \alpha \ln(t) - \alpha \ln(\beta) \quad (3.2)$$

โดยที่ $\ln\left(\ln\left(\frac{1}{1-F(t)}\right)\right)$ เป็นค่าตามแกน Y, $\ln(t)$ เป็นค่าตามแกน X, α จะมีค่าเท่ากับ

ความชันของกราฟ และ $-\alpha \ln(\beta)$ คือระยะตัดแกน Y

ในการวิเคราะห์ความน่าจะเป็นสะสมของความผิดปกติของอุปกรณ์ด้วยการแจกแจงไวบูลนั้น ตัวแปรของขนาดหรือ β จะมีค่าเท่ากับอายุการใช้งานเฉลี่ยของอุปกรณ์ และตัวแปรรูปร่างหรือ α จะแสดงถึงรูปแบบของความผิดปกติ ซึ่งในงานวิจัยนี้จะพิจารณาความผิดปกติด้วยการแจกแจงไวบูลเฉพาะ $\alpha > 1$ หรืออัตราความผิดปกติเพิ่มขึ้นตามช่วงเวลา ซึ่งเป็นช่วงที่การประมาณค่าด้วยการแจกแจงไวบูลมีประสิทธิภาพสูงสุด

3.1.3 ทำการเปรียบเทียบความเหมาะสมของข้อมูลจาก 3.1.2 กับการแจกแจงไวบูลในรูปแบบต่างๆ ได้แก่

1. การแจกแจงไวบูล (Hoang, 2006)

วิธีนี้จะคำนวณหาการแจกแจงความผิดปกติของเครื่องจักรที่เกิดจากรวมรวมเหตุการณ์ความผิดปกติทั้งหมดที่ได้จาก 3.1.2 นำมาประมาณค่าการแจกแจงไวบูลรวมทุกเหตุการณ์ซึ่งมีฟังก์ชันความน่าจะเป็นสะสมดังนี้

$$F(t; \alpha, \beta) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha} \quad (3.3)$$

2. การแจกแจงไวบูลเชิงประกอบ (Roy et al. 2021)

วิธีนี้จะคำนวณหาการแจกแจงความผิดปกติของเครื่องจักร จากการนำการแจกแจงไวบูลของแต่ละเหตุการณ์ความผิดปกติมาประกอบกัน กล่าวคือแต่ละเหตุการณ์จะมีการแจกแจงไวบูลที่ต่างกัน ซึ่งมีฟังก์ชันความน่าจะเป็นสะสมดังนี้

$$F(X) = \begin{cases} F_{t_0}(X) & \text{เมื่อ } X \text{ อยู่ในช่วงเหตุการณ์ที่ } 1 \\ F_{t_{n-1}}(X) & \text{เมื่อ } X \text{ อยู่ในช่วงเหตุการณ์ที่ } n \end{cases} \quad (3.4)$$

โดยที่ t คือจุดแบ่งกลุ่มการแจกแจงไวบูลที่ได้จากการแบ่งกลุ่มด้วยข้อมูลของเหตุการณ์ความผิดปกติในแต่ละเหตุการณ์ ในงานวิจัยของ Roy et al (2021) ได้ใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนมาช่วยในการหาจุด t ที่เหมาะสมของ 2 เหตุการณ์ แต่ในงานวิจัยนี้มีจำนวนเหตุการณ์มากกว่า 2 เหตุการณ์ จึงจะใช้วิธีต้นไม้ตัดสินใจแทนเพื่อลดการซับซ้อนในการคำนวณที่ละคู่ของวิธีซัพพอร์ตเวกเตอร์แมชชีน

3. การแจกแจงไวบูลผสมเชิงเส้น (Roy et al. 2021)

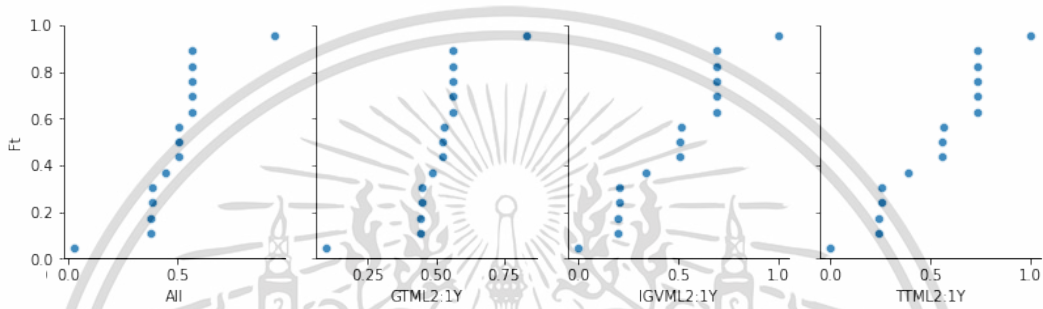
วิธีนี้พัฒนามาจากวิธีการแจกแจงไวบูลเชิงประกอบ ซึ่งจะผสมการแจกแจงไวบูลในแต่ละเหตุการณ์ความผิดปกติที่มีค่าสัมประสิทธิ์การถดถอยต่างกันในแต่ละช่วงเหตุการณ์ ฟังก์ชันความน่าจะเป็นสะสมมีดังนี้

$$F(t) = \sum_{i=1}^n w_{i,t} F_i(t) \quad (3.5)$$

โดยที่ $\sum_{i=1}^n w_{i,t} = 1$, w_i คือค่าสัมประสิทธิ์การถดถอยของการแจกแจงที่เหตุการณ์ที่ i และ t คือจุดแบ่งกลุ่มการแจกแจง $F_i(t)$ ที่ได้จากวิธีการแจกแจงไวบูลในแต่ละช่วงเหตุการณ์

4. การแจกแจงไวบูลผสมพหุนาม (Polynomial Mixed Weibull Distribution)

ในงานวิจัยนี้เสนอการแจกแจงไวบูลผสมพหุนาม เป็นการรวมการแจกแจงไวบูลในแต่ละเหตุการณ์ความผิดปกติทั้งหมด ซึ่งจากการตรวจสอบข้อมูลพบว่าจากกราฟแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ $F_i(t)$ และตัวแปรตาม $F(t)$ มีความสัมพันธ์แบบพหุนามหลายตัวแปรดังรูป



รูปที่ 3.1 แผนภูมิแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ $F_i(t)$ และตัวแปรตาม $F(t)$

จากรูปที่ 3.1 จะเห็นว่าตัวแปรอิสระ $F_i(t)$ หรือความน่าจะเป็นสะสมของแต่ละรูปแบบความผิดปกติมีความสัมพันธ์ไม่เป็นเชิงเส้นกับตัวแปรตาม $F(t)$ หรือความน่าจะเป็นสะสมที่ได้จากการประมาณค่าของปีนาร์ด จึงได้นำสมการพหุนามหลายตัวแปรมาใช้ในการประมาณค่าการแจกแจงดังสมการ

$$F(t, \beta) = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_1 F_2 + \beta_4 F_1^2 + \beta_5 F_2^2 + \dots \quad (3.6)$$

ในการประมาณค่า β ในสมการถดถอยแบบพหุนามหลายตัวแปร จะใช้วิธีกำลังสองน้อยสุด (Least Square Method)

$$\text{residual}_i = Y_i - F(t_i, \beta) \quad (3.7)$$

$$\beta = \beta | \min \sum_{i=1}^n \text{residual}_i^2 \quad (3.8)$$

ในงานวิจัยนี้ได้ทดลองหาอันดับของพหุนามที่เหมาะสมที่น้อยที่สุดจากค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้วของตัวแปรอิสระที่สัมพันธ์กับตัวแปรตามและค่าการแจกแจงปกติของส่วนเหลือ

3.1.4 ทำการเปรียบเทียบประสิทธิภาพของแต่ละการแจกแจงสะสม $F(t)$ ใน 3.1.3 กับ ข้อมูลความน่าจะเป็นของความผิดปกติในข้อ 3.1.2 ด้วยความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย โดยพิจารณาจากค่าที่ต่ำสุด

3.1.5 หาค่าความน่าจะเป็นของความผิดปกติจากการแจกแจงสะสม $F(t)$ ที่ได้จากข้อ 3.1.4 โดยที่ t คือจำนวนวันที่นับจากเครื่องจักรพร้อมใช้งานล่าสุด

$$P_{\text{maintenance}} = F(t) \quad (3.10)$$

3.2 ข้อมูลการใช้งานเครื่องจักร (Operation Data)

ในงานวิจัยนี้ได้พัฒนาตัวแบบจากข้อมูลการใช้งานเครื่องจักรในสภาวะปกติจากตัวรับสัญญาณประเภทต่างๆ เช่น ค่าสัญญาณการสั่น ค่าสัญญาณอุณหภูมิ ค่าสัญญาณความดัน ค่าสัญญาณอัตราการไหล และค่าสัญญาณความเร็วรอบของเครื่องจักร เป็นต้น ตามขั้นตอนต่อไปนี้

3.2.1 กำหนดขนาดตัวอย่างข้อมูลการใช้งานของเครื่องจักรในสถานีเพิ่มความดันก๊าซธรรมชาติอย่างน้อย 7 วัน วันละ 24 ชั่วโมง = 168 ตัวอย่างต่ออุปกรณ์ จำนวน 100 อุปกรณ์

3.2.2 แบ่งกลุ่มข้อมูลเป็น 2 ส่วน โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลกับความเร็วยรอบของเครื่องจักร (correlation, R) และพิจารณาจำนวนข้อมูล (n) อ้างอิงงานวิจัย Guyon, 2003

$$R = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (3.11)$$

$$t = R\sqrt{n}$$

$t < 3$ หมายถึงข้อมูลที่ไม่มีความสัมพันธ์กับความเร็วยรอบของเครื่องจักร จะใช้ในการกำหนดขีดจำกัดควบคุมในข้อ 3.2.3 และ $t > 3$ หมายถึงข้อมูลที่มีความสัมพันธ์กับความเร็วยรอบของเครื่องจักร จะใช้ในการกำหนดขีดจำกัดควบคุมสองตัวแปรในข้อ 3.2.4

3.2.3 ข้อมูลการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วยรอบของเครื่องจักร

3.2.3.1 เตรียมข้อมูลการใช้งานเครื่องจักรโดยวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ให้อยู่ในรูปแบบการแจกแจงปกติ และตรวจสอบด้วยการทดสอบจาร์คูบีร่า (Jarque-Bera test) ซึ่งเป็นการทดสอบภาวะสารูปดี (Goodness of fit test) ถ้ามีการแจกแจงปกติ ค่าพีจะต้องมากกว่าระดับนัยสำคัญที่กำหนด (0.05) (Das, 2016)

3.2.3.2 วิธีการหาค่าความผิดปกติ

1. วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว (Univariate k-sigma)

วิธีนี้จะกำหนดขีดจำกัดควบคุมที่ยอมรับได้เท่ากับ k เท่าของส่วนเบี่ยงเบนมาตรฐานของประชากร เนื่องจากเราไม่สามารถหาส่วนเบี่ยงเบนมาตรฐานของประชากรที่แท้จริงได้ จึงใช้ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่างแทน (Cetin and Tasgin, 2021)

$$S.D. = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n-1}}$$

เมื่อ Y เป็นตัวแปรสุ่มของข้อมูลการเดินเครื่องในแต่ละอุปกรณ์ที่ไม่สัมพันธ์กับข้อมูลความเร็วรอบของเครื่องจักร

n เป็นขนาดข้อมูลที่ใช้พิจารณา

สามารถกำหนดขีดจำกัดควบคุมดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \bar{Y} + 3 S.D.$$

$$\text{เส้นกึ่งกลาง} = \text{ค่าเฉลี่ย } \bar{Y}$$

$$\text{ขีดจำกัดควบคุมล่าง} = \bar{Y} - 3 S.D.$$

และสามารถหาค่าความผิดปกติได้จากค่าคะแนนมาตรฐาน

$$Z_i = \frac{Y_i - \bar{Y}}{S.D.} \quad (3.12)$$

2. วิธีป่าไม้โดดเดี่ยว (Isolation Forest)

วิธีนี้จะนำข้อมูลของตัวแปรสุ่มมาแบ่งกลุ่มด้วยการลดทอนความสัมพันธ์ระหว่างข้อมูล ดังนั้นเมื่อมีค่าสังเกตเพิ่มเข้ามา จะพิจารณาค่าความสัมพันธ์กับข้อมูลฝึก หากมีค่าความสัมพันธ์ต่ำจะได้ผลลัพธ์การทำนายว่าเป็นค่าสังเกตที่ผิดปกติ สามารถกำหนดพารามิเตอร์ที่ส่งผลต่อแบบจำลองดังนี้ ค่าการปนเปื้อนเป็นตัวกำหนดสัดส่วนของความผิดปกติของชุดข้อมูล จำนวนคุณสมบัติสูงสุดที่ใช้ในการฝึกแบบจำลองหากมีค่าสูงจะส่งผลให้ค่าความสัมพันธ์ระหว่างข้อมูลเพิ่มขึ้นด้วย จำนวนข้อมูลสูงสุดในชุดข้อมูลฝึกและ จำนวนต้นไม้หากมีค่าสูงจะส่งผลต่อการเกิดปัญหาตัวแบบจดจำข้อมูลฝึกเท่านั้น ทำให้ค่าเรียกคืนในชุดข้อมูลทดสอบต่ำ (Overfitting) ในตัวแบบนี้สามารถหาค่าความผิดปกติได้จากจำนวนโหนดตัดสินใจเฉลี่ยของป่าสุ่มได้ (Fei et al., 2008)

3.2.4 ข้อมูลการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วยรอบของเครื่องจักร

3.2.4.1 เตรียมข้อมูลและตรวจสอบสมมติฐานของสมการถดถอยเชิงเส้นดังนี้

1. ตรวจสอบข้อมูลการใช้งานเครื่องจักรมีความสัมพันธ์กับความเร็วยรอบของเครื่องจักรเชิงเส้น จากค่าสัมประสิทธิ์สหสัมพันธ์กำหนดให้ต้องมีค่าสหสัมพันธ์มากกว่า 0.7

2. ตรวจสอบส่วนเหลือที่ได้จากสมการถดถอยเชิงเส้นมีการแจกแจงปกติด้วยการทดสอบ] กำหนดให้ระดับนัยสำคัญเท่ากับ 0.05 กล่าวคือต้องมีค่าทดสอบพีมากกว่า 0.05

3. ตรวจสอบความเป็นอิสระต่อกันระหว่างข้อมูลการใช้งานเครื่องจักรและส่วนเหลือที่ได้จากสมการถดถอยเชิงเส้นด้วยการทดสอบเดอร์บินวัตสัน กำหนดให้ค่าเดอร์บินวัตสันอยู่ในช่วง 1.5 – 2.5

4. ความแปรปรวนของส่วนเหลือที่ได้จากสมการถดถอยเชิงเส้นคงที่ด้วยการทดสอบไวกซ์ กำหนดให้ระดับนัยสำคัญเท่ากับ 0.05 กล่าวคือต้องมีค่าทดสอบพีมากกว่า 0.05

ในงานวิจัยนี้จะพิจารณาตัวแบบที่เป็นไปตามสมมติฐานทั้ง 4 ข้อเพื่อนำไปใช้เปรียบเทียบวิธีการหาความผิดปกติต่อไป

3.2.4.2 วิธีการหาค่าความผิดปกติ

1. วิธีแผนภูมิควบคุมถดถอย (Regression control chart)

วิธีนี้จะพิจารณาความคลาดเคลื่อนมาตรฐาน ร่วมกับค่าสัมประสิทธิ์การกำหนดพหุคูณระหว่างข้อมูลการใช้งานเครื่องจักรในแต่ละตัวตรวจจับสัญญาณและความเร็วยรอบของเครื่องจักร เนื่องจากเราไม่สามารถหาส่วนเบี่ยงเบนมาตรฐานของประชากรที่แท้จริงได้ จึงใช้ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่างแทน (Hayati, 2017)

$$S.D._y = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}}$$

และสามารถหาค่าสัมประสิทธิ์การตัดสินใจพหุคูณได้จาก

$$r^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \quad (3.13)$$

$$\hat{Y}_i = \beta_0 + \beta_1 X_i$$

โดยที่
$$\beta_0 = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2}$$

และ
$$\beta_1 = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ Y เป็นตัวแปรสุ่มของข้อมูลการเดินเครื่องในแต่ละตัวตรวจจับสัญญาณ ที่สัมพันธ์กับข้อมูลความเร็วรอบของเครื่องจักรแทนด้วย X

n เป็นจำนวนข้อมูลที่ใช้พิจารณา

สามารถกำหนดขีดจำกัดควบคุมถอยดังนี้

$$S_{error} = S.D. \sqrt{1-r^2} \quad (3.14)$$

$$\text{ขีดจำกัดควบคุมบน} = \hat{Y}_i + 3 S_{error}$$

$$\text{เส้นกึ่งกลาง} = \text{ค่าทำนาย } \hat{Y}_i$$

$$\text{ขีดจำกัดควบคุมล่าง} = \hat{Y}_i - 3 S_{error}$$

และสามารถหาค่าความผิดปกติได้จากค่าคะแนนมาตรฐาน

$$Z_i = \frac{Y_i - \hat{Y}_i}{S_{error}} \quad (3.15)$$

2) วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร (Multi-variate k -sigma)

วิธีนี้จะพิจารณาระยะทางของเวกเตอร์เทียบกับเวกเตอร์เฉลี่ยของการกระจายตัวของจุดข้อมูล ในปริภูมิของ 2 มิติ ประกอบด้วยข้อมูลการเดินเครื่องในแต่ละตัวตรวจจับสัญญาณและข้อมูลความเร็วรอบของเครื่องจักร สามารถคำนวณได้จากระยะทางมาฮาလာโนบิส (Cetin and Tasgin, 2021)

$$k_i = \sqrt{(\bar{X}_i - \bar{\pi})^T \Sigma^{-1} (\bar{X}_i - \bar{\pi})} \quad (3.16)$$

โดยที่ \bar{X}_i คือเวกเตอร์ของจุดข้อมูลในตัวแปรสุ่ม

$\bar{\pi}$ คือเวกเตอร์ค่าเฉลี่ยของจุดข้อมูลในตัวแปรสุ่ม

Σ คือเมทริกซ์ความแปรปรวนร่วมของจุดข้อมูล

ซึ่งจะสามารถหาค่าความผิดปกติได้จากค่า k หากค่า k สูงจะมีโอกาสความผิดปกติสูงด้วย

3. วิธี k เท่าของส่วนเหลือมาตรฐาน (k standard residual)

ในงานวิจัยนี้เสนอการหาขีดจำกัดควบคุมจากค่าส่วนเหลือมาตรฐาน ซึ่งพิจารณาได้จากการแจกแจงของส่วนเหลือของตัวแบบ หากค่าสังเกตมีส่วนเหลือที่มีค่าสูงกว่าค่ามาตรฐานมากจะมีแนวโน้มความผิดปกติสูงด้วย นอกจากนี้วิธีดังกล่าวสามารถใช้ทำนายค่าความผิดปกติได้ทั้งตัวแปรเดียวและหลายตัวแปร แต่การแจกแจงของส่วนเหลือของตัวแบบนี้ต้องมีการแจกแจงแบบปกติสามารถกำหนดขีดจำกัดควบคุมได้ดังนี้

$$S_{residual} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ในงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{ขีดจำกัดควบคุมบน} = \hat{Y}_i + 3 S_{residual}$$

$$\text{เส้นกึ่งกลาง} = \text{ค่าทำนาย } \hat{Y}_i$$

$$\text{ขีดจำกัดควบคุมล่าง} = \hat{Y}_i - 3 S_{residual}$$

และสามารถหาค่าความผิดปกติได้จากค่าคะแนนมาตรฐาน

$$Z_i = \sqrt{\frac{Y_i - \hat{Y}_i}{S_{residual}}} \quad (3.17)$$

3.2.5 การเปรียบเทียบประสิทธิภาพการตรวจจับความผิดปกติ

ข้อมูลการใช้งานเครื่องจักรที่นำมาคำนวณตัวแบบสำหรับหาค่าความผิดปกติของข้อมูลนี้อาจจะเกิดขึ้นเป็นข้อมูลการใช้งานเครื่องจักรในสภาวะปกติทั้งหมด ดังนั้นในการทดสอบประสิทธิภาพของตัวแบบนี้ จะสุ่มตัวอย่างข้อมูลการใช้งานในสภาวะปกติจำนวน 100 ตัวอย่าง และเพิ่มข้อมูลสังเคราะห์ข้อมูลการใช้งานสภาวะผิดปกติจำนวน 100 ตัวอย่างที่มีการแจกแจงปรกติ รวมทั้งสิ้น 200 ตัวอย่าง เพื่อนำไปวัดประสิทธิภาพด้วยค่าความเที่ยง ค่าเรียกคืน และคะแนน F1

3.2.6 การหาค่าความน่าจะเป็นของความผิดปกติ

ในงานวิจัยนี้หาค่าความน่าจะเป็นของความผิดปกติจากค่าคะแนนมาตรฐานที่ได้จากวิธีที่มีคะแนน F1 ค่าเรียกคืน และค่าความเที่ยงสูงสุดตามในข้อ 3.2.6 (Cetin and Tasgin 2021)

$$P_{operation} = \begin{cases} s(k_i - c), k_i = |z| \geq c \\ 0, otherwise \end{cases} \quad (3.18)$$

เมื่อ s คือความน่าจะเป็นสะสมของการแจกแจงปรกติ

Z คือคะแนนมาตรฐาน

c เป็นค่าคงที่ของช่วงการเบี่ยงเบนจากค่าเฉลี่ยที่ยอมรับได้

3.3 การหาค่าความผิดปกติรวม

การคำนวณค่าร้อยละความผิดปกติรวมในงานวิจัยนี้เนื่องจากข้อมูลบำรุงรักษาเครื่องจักรและข้อมูลการใช้งานเครื่องจักรเป็นอิสระต่อกัน จึงสามารถคำนวณจากผลคูณค่าความน่าจะเป็นจากข้อมูลการบำรุงรักษาเครื่องจักรและข้อมูลการใช้งานเครื่องจักรได้ดังสมการ (Yenal,2020)

$$\% Failure = P_{maintenance} \times P_{operation} \times 100$$

ในงานวิจัยครั้งนี้กำหนดร้อยละความผิดปกติรวมที่ยอมรับได้อยู่ที่ 20% หากมีค่ามากกว่า 20% จะดำเนินการบันทึกผลและแจ้งเตือนเพื่อหาสาเหตุของความผิดปกติ เพื่อใช้เป็นข้อมูลในการวางแผนงานบำรุงรักษาหรือการประเมินความเสี่ยงก่อนใช้งานเครื่องจักร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

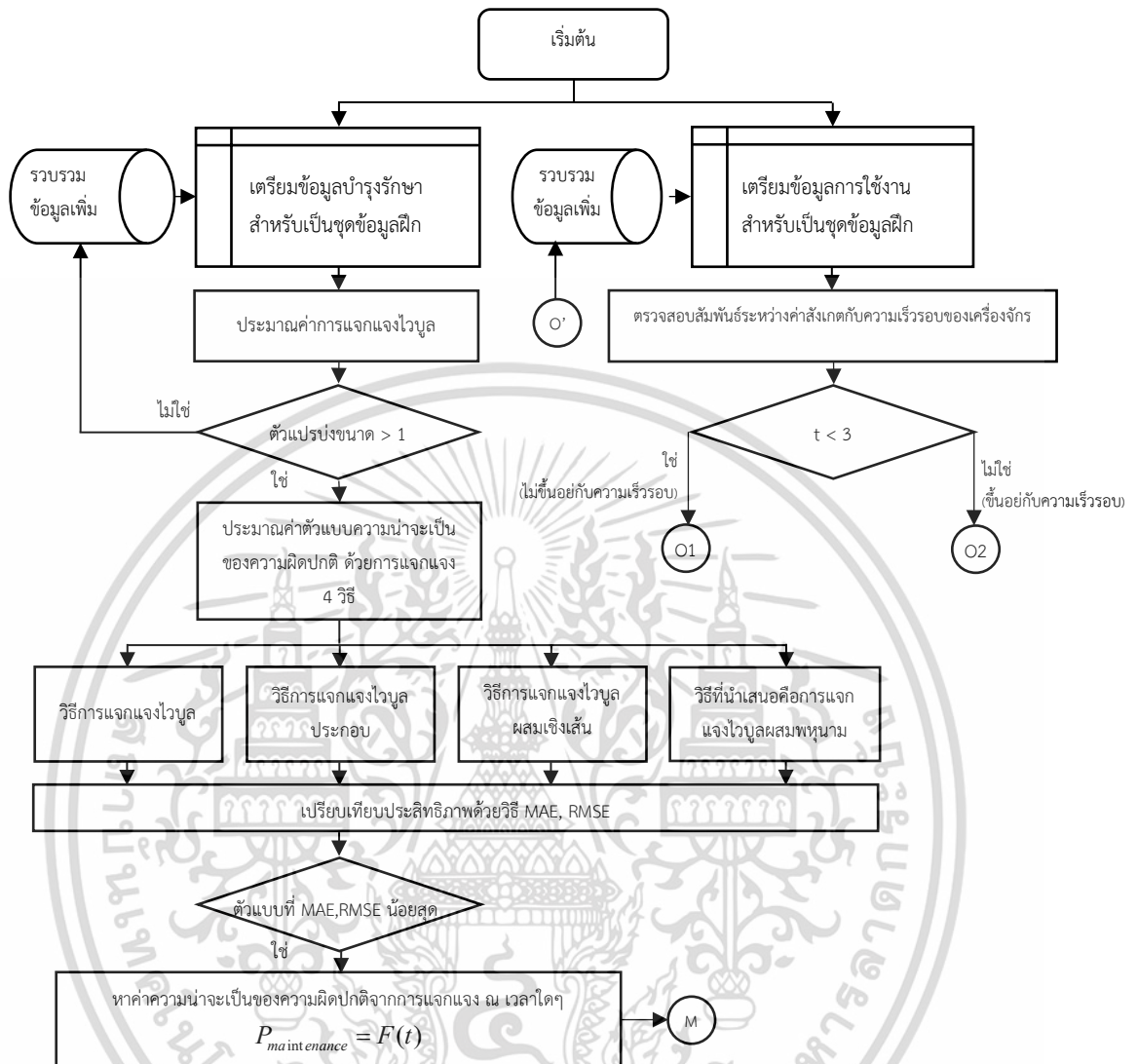
3.4 เครื่องมือที่ใช้ในงานวิจัย

- 3.4.1 สิ่งแวดล้อมสำหรับการพัฒนาแบบเบ็ดเสร็จ (IDE) Jupyter notebooks
- 3.4.2 โปรแกรมภาษาไพธอน เวอร์ชัน 3.9.12
- 3.4.3 ไลบรารีในการเตรียมข้อมูล เช่น pandas, numpy, math
- 3.4.4 ไลบรารีในการทำแบบจำลอง เช่น statmodel, sklearn
- 3.4.5 ไลบรารีในการแสดงข้อมูล เช่น matplotlib, seaborn
- 3.4.6 ไลบรารีในการบันทึกและเชื่อมต่อฐานข้อมูล เช่น pickle, schemy
- 3.4.7 ระบบจัดการฐานข้อมูล Microsoft SQL server 2014



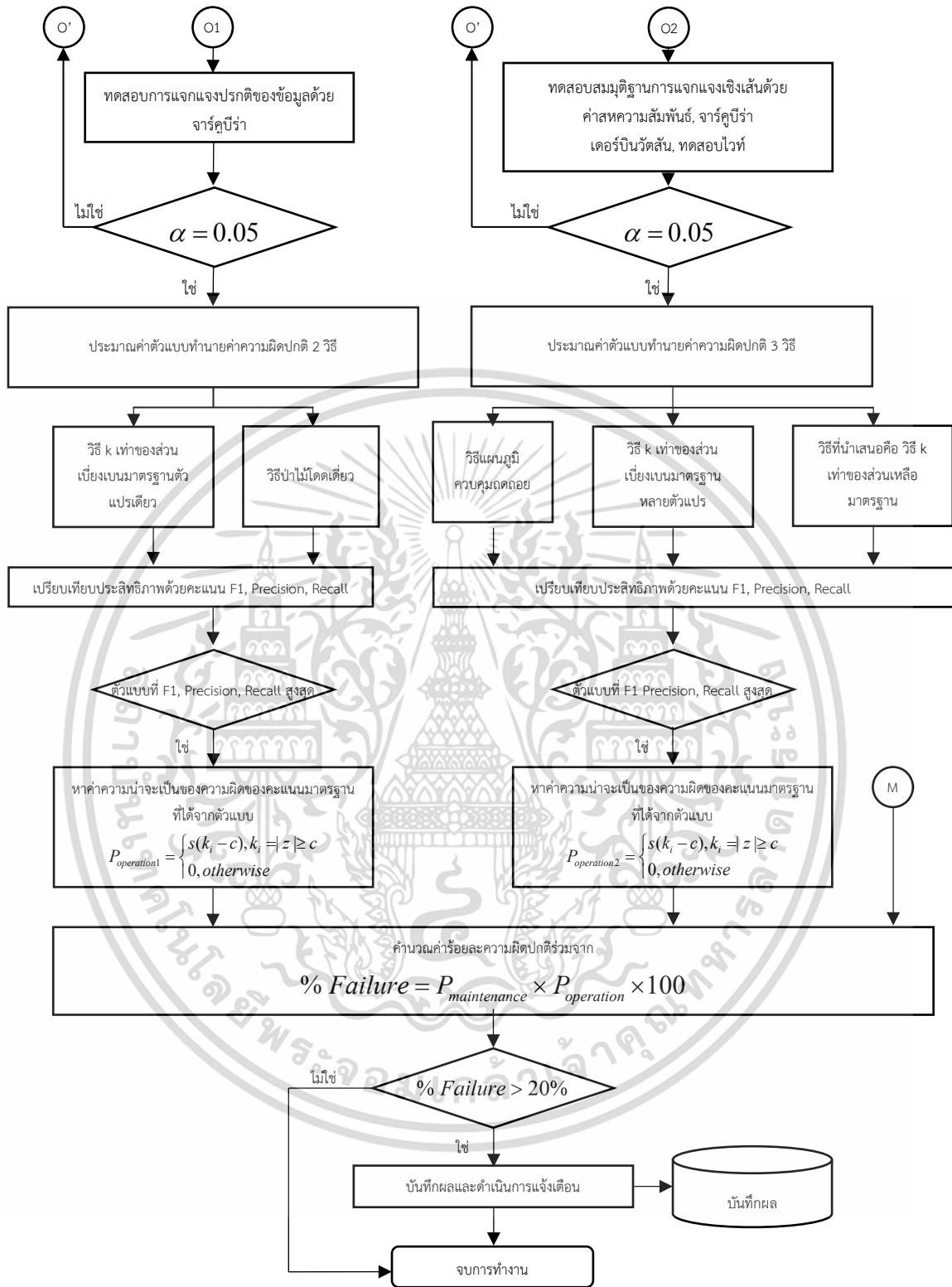
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 สรุปขั้นตอนการดำเนินงาน



รูปที่ 3.2 ขั้นตอนการทำงานเพื่อคำนวณค่าความผิดปกติของเครื่องจักร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 ขั้นตอนการทำงานเพื่อคำนวณค่าความผิดพลาดของเครื่องจักร (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

ในงานวิจัยนี้ศึกษาเกี่ยวกับการคำนวณค่าความน่าจะเป็นของความผิดปกติของเครื่องจักรในสถานีเพิ่มความดันก๊าซธรรมชาติ มีวัตถุประสงค์ในการนำข้อมูลการบำรุงรักษาเครื่องจักรและการใช้งานเครื่องจักรมาบูรณาการร่วมกับการวิเคราะห์ทางสถิติและการเรียนรู้ของเครื่อง เพื่อเพิ่มประสิทธิภาพในการคำนวณค่าความน่าจะเป็นของความผิดปกติของเครื่องจักร

4.1 ผลการวิจัยข้อมูลการบำรุงรักษาเครื่องจักร

4.1.1 การเตรียมและตรวจสอบข้อมูล

ในงานวิจัยนี้จะพิจารณาอายุการใช้งานเครื่องจักร (Time to fail) แบ่งตามแต่ละรูปแบบความผิดปกติ (Failure Mode) ซึ่งจากทั้งหมด 119 ตัวอย่างสามารถหาอายุการใช้งานเครื่องจักรได้ 3 รูปแบบความผิดปกติ ดังตารางที่ 4.1 กล่าวคือรูปแบบความผิดปกติที่เกิดขึ้นครั้งเดียวจะไม่สามารถหาอายุการใช้งานของเครื่องจักรได้

ตารางที่ 4.1 อายุการใช้งานเครื่องจักรของแต่ละรูปแบบความผิดปกติ

เครื่องจักร	วันที่เครื่องจักรพบความผิดปกติ	รูปแบบความผิดปกติ	อายุการใช้งานเครื่องจักร
TSO-ENGIN_C	1/11/2021	GTML2:1Y	61
TSO-ENGIN_C	1/9/2021	GTML2:1Y	304
TSO-ENGIN_A	1/10/2021	GTML2:1Y	334
TSO-ENGIN_C	1/10/2022	GTML2:1Y	334
TSO-ENGIN_A	1/9/2022	GTML2:1Y	335
TSO-ENGIN_B	1/10/2022	GTML2:1Y	699
TSO-ENGIN_C	1/3/2021	IGVML2:1Y	273
TSO-ENGIN_B	1/2/2021	IGVML2:1Y	276
TSO-ENGIN_A	1/1/2022	IGVML2:1Y	365
TSO-ENGIN_B	1/2/2022	IGVML2:1Y	365
TSO-ENGIN_C	1/3/2022	IGVML2:1Y	365
TSO-ENGIN_C	1/3/2021	TTML2:1Y	273
TSO-ENGIN_B	1/2/2021	TTML2:1Y	276
TSO-ENGIN_A	1/1/2022	TTML2:1Y	365
TSO-ENGIN_B	1/2/2022	TTML2:1Y	365

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.1 แสดงอายุการใช้งานของเครื่องจักรจากความผิดปกติแต่ละรูปแบบ ซึ่งจะใช้ในการคำนวณความน่าจะเป็นของความผิดปกติของเครื่องจักรจากการหาค่าอันดับที่ของมัธยฐานหรือการประมาณค่าความน่าจะเป็นสะสมด้วยวิธีการประมาณค่าของปีนาร์ด

ตารางที่ 4.2 ความน่าจะเป็นสะสมของความผิดปกติในแต่ละรูปแบบ

เครื่องจักร	วันที่เครื่องจักรพบความผิดปกติ	รูปแบบความผิดปกติ	อายุการใช้งาน	ความน่าจะเป็นสะสมของความผิดปกติ
TSO-ENGIN_C	1/11/2021	GTML2:1Y	61	0.109375
TSO-ENGIN_C	1/9/2021	GTML2:1Y	304	0.265625
TSO-ENGIN_A	1/10/2021	GTML2:1Y	334	0.421875
TSO-ENGIN_C	1/10/2022	GTML2:1Y	334	0.578125
TSO-ENGIN_A	1/9/2022	GTML2:1Y	335	0.734375
TSO-ENGIN_B	1/10/2022	GTML2:1Y	699	0.890625
TSO-ENGIN_C	1/3/2021	IGVML2:1Y	273	0.12963
TSO-ENGIN_B	1/2/2021	IGVML2:1Y	276	0.314815
TSO-ENGIN_A	1/1/2022	IGVML2:1Y	365	0.5
TSO-ENGIN_B	1/2/2022	IGVML2:1Y	365	0.685185
TSO-ENGIN_C	1/3/2022	IGVML2:1Y	365	0.87037
TSO-ENGIN_C	1/3/2021	TTML2:1Y	273	0.159091
TSO-ENGIN_B	1/2/2021	TTML2:1Y	276	0.386364
TSO-ENGIN_A	1/1/2022	TTML2:1Y	365	0.613636
TSO-ENGIN_B	1/2/2022	TTML2:1Y	365	0.840909

จากตารางที่ 4.2 แสดงความน่าจะเป็นสะสมของความผิดปกติในแต่ละรูปแบบซึ่งจะถูกนำมาทดสอบการแจกแจงไวบูล กล่าวคือการประมาณค่าพารามิเตอร์การแจกแจงไวบูล ซึ่งประกอบด้วยตัวแปรรูปร่าง (α) และตัวแปรบ่งขนาด (β) ในงานวิจัยนี้จะประมาณค่าด้วยวิธี Curve fitting และจะพิจารณาแบบความผิดปกติที่มีตัวแปรรูปร่างมากกว่า 1 หรือมีอัตราความผิดปกติเพิ่มขึ้นตามช่วงเวลาเท่านั้น

ตารางที่ 4.3 พารามิเตอร์การแจกแจงไวบูลในแต่ละรูปแบบความผิดปกติ

รูปแบบความผิดปกติ	พารามิเตอร์การแจกแจงไวบูล $[\alpha, \beta]$
GTML2:1Y	[1.1765851769210385, 428.8511415825653]
IGVML2:1Y	[5.80010819318955, 354.4428254941219]
TTML2:1Y	[5.379088446447412, 345.9295764141948]

จากตารางที่ 4.3 แสดงค่าตัวแปรรูปร่าง (α) และตัวแปรขนาด (β) ซึ่งจะเห็นว่าทุกรูปแบบความผิดปกติมีค่าพารามิเตอร์ตัวแปรรูปร่างมากกว่า 1 จึงจะใช้ในการพัฒนาตัวแบบต่อไป ทั้งนี้ในงานวิจัยนี้จะพัฒนาตัวแบบจากข้อมูลการบำรุงรักษาเครื่องจักรทุกรูปแบบความผิดปกติร่วมกัน เพื่อให้สามารถนำไปใช้งานได้ง่ายจำนวน 4 วิธีได้แก่ วิธีการแจกแจงไวบูล วิธีการแจกแจงไวบูลเชิงประกอบ วิธีการแจกแจงไวบูลผสมเชิงเส้น และการแจกแจงไวบูลผสมพหุนาม

4.1.2 การพัฒนาตัวแบบจากข้อมูลการบำรุงรักษาเครื่องจักร

1) วิธีการแจกแจงไวบูล (Weibull distribution)

ทำการหาค่าความน่าจะเป็นสะสมของความผิดปกติของเครื่องจักรจากทุกรูปแบบความผิดปกติร่วมกันผ่านการประมาณค่าการแจกแจงไวบูล

$$F(t; \alpha, \beta) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha}$$

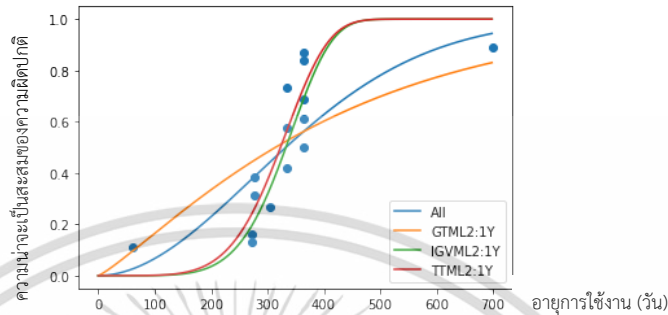
ตารางที่ 4.4 ความน่าจะเป็นสะสมของความผิดปกติของเครื่องจักร

เครื่องจักร	วันที่เครื่องจักรพบความผิดปกติ	อายุการใช้งานเครื่องจักร	ความน่าจะเป็นสะสมของความผิดปกติ
TSO-ENGIN_C	1/11/2021	61	0.045455
TSO-ENGIN_C	1/3/2021	273	0.11039
TSO-ENGIN_C	1/3/2021	273	0.175325
TSO-ENGIN_B	1/2/2021	276	0.24026
TSO-ENGIN_B	1/2/2021	276	0.305195
TSO-ENGIN_C	1/9/2021	304	0.37013
TSO-ENGIN_A	1/10/2021	334	0.435065
TSO-ENGIN_C	1/10/2022	334	0.5
TSO-ENGIN_A	1/9/2022	335	0.564935
TSO-ENGIN_A	1/1/2022	365	0.62987
TSO-ENGIN_B	1/2/2022	365	0.694805
TSO-ENGIN_C	1/3/2022	365	0.75974
TSO-ENGIN_A	1/1/2022	365	0.824675
TSO-ENGIN_B	1/2/2022	365	0.88961
TSO-ENGIN_B	1/10/2022	699	0.954545

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเฉพาะที่ศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้เพื่อเผยแพร่ในด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.4 แสดงข้อมูลอายุการใช้งานของเครื่องจักรโดยไม่แบ่งตามรูปแบบความผิดปกติ สามารถประมาณค่าพารามิเตอร์การแจกแจงไวบูลล์ด้วยวิธีของบิแนร์ดพบว่าตัวแปรปรั้งเท่ากับ 1.92 และตัวแปรบ่งขนาดเท่ากับ 402.44 การแจกแจงไวบูลล์ที่ประมาณค่าได้จะมีลักษณะตามรูปที่ 4.1

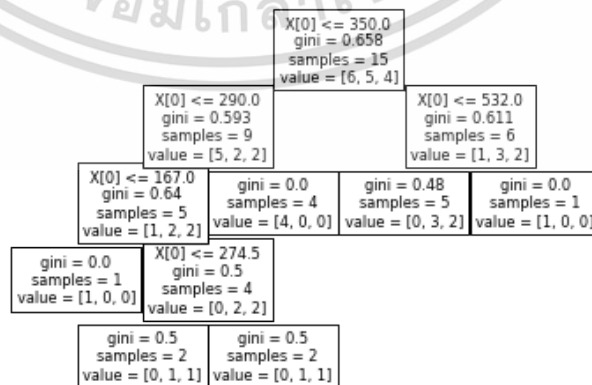


รูปที่ 4.1 แผนภูมิแสดงความน่าจะเป็นสะสมของความผิดปกติแต่ละรูปแบบ

จากรูปที่ 4.1 จะเห็นว่า การแจกแจงไวบูลล์ที่ข้อมูลอายุการใช้งานของเครื่องจักรไม่แบ่งตามรูปแบบความผิดปกติที่แสดงในเส้นสีน้ำเงินนั้น สามารถประมาณค่าความน่าจะเป็นสะสมของความผิดปกติในทุกจุดข้อมูลได้ และในวิธีถัดไปจะทดลองนำการแจกแจงไวบูลล์ที่ข้อมูลอายุการใช้งานของเครื่องจักรแบ่งตามรูปแบบความผิดปกติที่แสดงในเส้นสีเหลือง สีเขียว และสีแดงนั้น มาพิจารณา ร่วมกันเพื่อหาความน่าจะเป็นสะสมของความผิดปกติ

2) วิธีการแจกแจงไวบูลล์เชิงประกอบ (Composite Weibull distribution)

วิธีนี้จะคำนวณหาความน่าจะเป็นสะสมของความผิดปกติทุกรูปแบบ จากการนำการแจกแจงไวบูลล์ของแต่ละรูปแบบความผิดปกติมาประกอบกัน โดยจะพิจารณาการแจกแจงไวบูลล์ของแต่ละรูปแบบความผิดปกติในแต่ละช่วงเวลาที่มีความคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยสุด ในการพิจารณาช่วงอายุการใช้งานที่เหมาะสมในงานวิจัยนี้ได้ใช้วิธีตัดตัดสินใจมาช่วยในการแบ่งกลุ่ม



รูปที่ 4.2 ต้นไม้ตัดสินใจการแบ่งกลุ่มช่วงอายุการใช้งานทุกรูปแบบความผิดปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.2 ต้นไม้ตัดสินใจที่โหนดรากสามารถแบ่งช่วงของอายุการใช้งานได้เป็น 2 ช่วงโดยมีจุดแบ่งอยู่ที่ 350 วัน หากพิจารณาค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยของแต่ละเหตุการณ์น้อยสุดในแต่ละช่วงดังตารางที่ 4.5

ตารางที่ 4.5 ความน่าจะเป็นสะสมของการแจกแจงไวบูลแต่ละรูปแบบความผิดพลาด

ช่วงของอายุการใช้งาน	อายุการใช้งาน	ความน่าจะเป็นสะสมของความผิดพลาด	การแจกแจงไวบูลรูปแบบความผิดพลาด GTML2:1Y	การแจกแจงไวบูลรูปแบบความผิดพลาด IGVML2:1Y	การแจกแจงไวบูลรูปแบบความผิดพลาด TTML2:1Y	การแจกแจงไวบูลเชิงประกอบ
ช่วง 1	61	0.045455	0.095886	0.000037	0.000088	0.000037
	273	0.11039	0.444445	0.197457	0.244088	0.197457
	273	0.175325	0.444445	0.197457	0.244088	0.197457
	276	0.24026	0.448655	0.208927	0.256788	0.208927
	276	0.305195	0.448655	0.208927	0.256788	0.208927
	304	0.37013	0.486796	0.336668	0.392902	0.336668
	334	0.435065	0.525358	0.507633	0.563066	0.507633
	334	0.5	0.525358	0.507633	0.563066	0.507633
	335	0.564935	0.526603	0.513697	0.568891	0.513697
ช่วง 2	365	0.62987	0.562738	0.694432	0.736744	0.736744
	365	0.694805	0.562738	0.694432	0.736744	0.736744
	365	0.75974	0.562738	0.694432	0.736744	0.736744
	365	0.824675	0.562738	0.694432	0.736744	0.736744
	365	0.88961	0.562738	0.694432	0.736744	0.736744
	699	0.954545	0.83082	1	1	1
	ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยช่วง 1			0.14179	0.049679	0.058950
ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยช่วง 2			0.18478	0.083519	0.076343	

จากตารางที่ 4.5 จะเห็นว่าในแต่ละช่วงของอายุการใช้งานนั้น วิธีการแจกแจงไวบูลเชิงประกอบจะมีค่าเท่ากับการแจกแจงไวบูลรูปแบบความผิดพลาด IGVML2:1Y ในช่วงอายุการใช้งานน้อยกว่า 350 วัน และมีค่าเท่ากับเท่ากับการแจกแจงไวบูลรูปแบบความผิดพลาด TTML2:1Y ในช่วงอายุการใช้งานมากกว่า 350 วัน

3) วิธีการแจกแจงไวบูลผสมเชิงเส้น (Linear Mixed Weibull distribution)

วิธีนี้จะพิจารณาค่าสัมประสิทธิ์การถดถอยของการแจกแจงไวบูลแต่ละรูปแบบความผิดพลาดในรูปแบบเชิงเส้น เพื่อให้ได้ค่าความน่าจะเป็นสะสมใกล้เคียงกับความน่าจะเป็นสะสมของความผิดพลาดที่มากที่สุด ในการคำนวณค่าสัมประสิทธิ์การถดถอยในงานวิจัยนี้ใช้วิธีกำลังสองน้อยสุดซึ่งได้ผลลัพธ์ดังตารางที่ 4.6

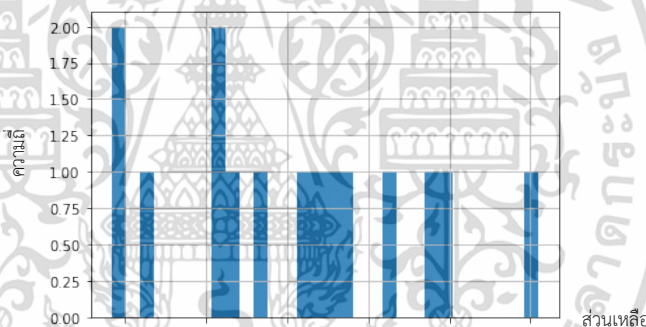
ตารางที่ 4.6 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นสะสมของความผิดปกติแต่ละรูปแบบ

รูปแบบความผิดปกติ	ค่าสัมประสิทธิ์การถดถอย
const	0.09
GTML2:1Y	-0.41
IGVML2:1Y	0.41
TTML2:1Y	0.81

Adj. R-squared: 0.918

จากตารางที่ 4.6 แสดงค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นที่มีค่าคงที่ในการคำนวณซึ่งสามารถเขียนสมการความน่าจะเป็นสะสมของความผิดปกติวิธีการแจกแจงไวบูลผสมเชิงเส้นได้เป็น

$$F(t) = 0.09 - 0.41(F_{GTML2:1Y}(t)) + 0.41(F_{IGVML2:1Y}(t)) + 0.81(F_{TTML2:1Y}(t))$$



รูปที่ 4.3 แผนภูมิแสดงความถี่ของส่วนเหลือที่ได้จากการแจกแจงไวบูลผสมเชิงเส้น

จากรูปที่ 4.3 แสดงการแจกแจงของส่วนเหลือที่ได้จากการแจกแจงไวบูลผสมเชิงเส้น ซึ่งตัวแบบที่เหมาะสมควรมีการแจกแจงปกติ

จากการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามพบว่ามีความสัมพันธ์แบบพหุนามดังรูปที่ 3.1 ทางผู้วิจัยจึงเสนอวิธีการแจกแจงไวบูลผสมพหุนาม

4) วิธีการแจกแจงไวบูลผสมพหุนาม (Polynomial Mixed Weibull Distribution)

ในงานวิจัยนี้เสนอการแจกแจงไวบูลผสมพหุนาม มีการคำนวณค่าสัมประสิทธิ์การถดถอยของการแจกแจงไวบูลแต่ละรูปแบบความผิดปกติในรูปแบบพหุนาม ในการทดลองหาอันดับของพหุนามที่เหมาะสมที่น้อยที่สุดจากค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้ว (Adjusted R-square) ของตัวแปรอิสระที่สัมพันธ์กับตัวแปรตามดังตารางที่ 4.7 และ 4.8 และค่าการแจกแจงปกติของส่วน

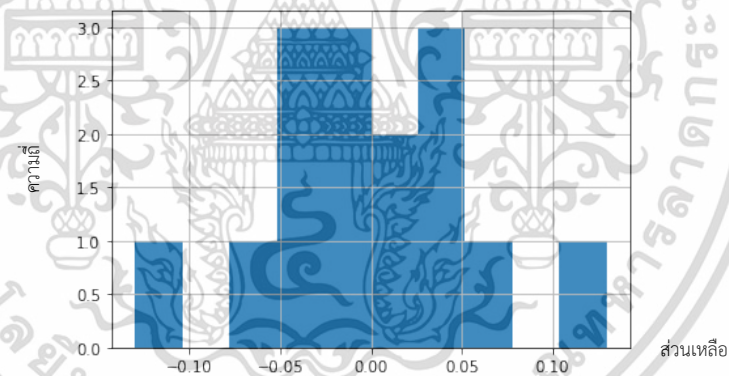
เหลือดังรูปที่ 4.7 และ 4.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นความผิดปกติสะสมแต่ละรูปแบบความผิดปกติ พหุนามอันดับ 2

รูปแบบความผิดปกติ	ค่าสัมประสิทธิ์การถดถอย
const	-996.0632
x1	1.40E+04
x2	-7.01E+04
x3	3.54E+04
x4	-3.77E+04
x5	-1.17E+05
x6	1.58E+05
x7	-3.88E+04
x8	1.03E+05
x9	-4.73E+04

Adj. R-squared: 0.918



Jarque Bera: (statistic=0.5049748115894388, pvalue=0.7768659967875418)

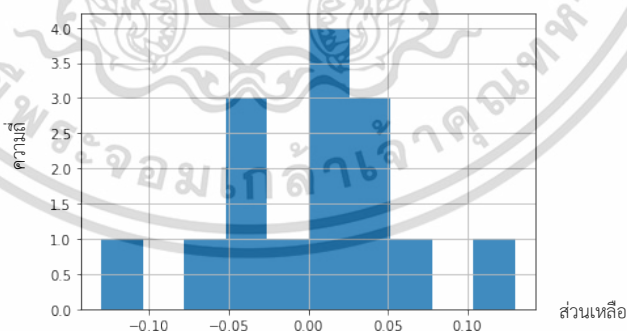
รูปที่ 4.4 แผนภูมิแสดงความผันแปรของส่วนเหลือที่ได้จากการแจกแจงไวบูลผสมพหุนามอันดับ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 ค่าสัมประสิทธิ์การถดถอยของความน่าจะเป็นความผิดปกติสะสมแต่ละรูปแบบความผิดปกติพหุนามอันดับ 3

รูปแบบความผิดปกติ	ค่าสัมประสิทธิ์การถดถอย
const	-220.9868
x1	2920.2073
x2	937.6963
x3	-2301.3514
x4	-6275.1335
x5	2921.403
x6	3219.7077
x7	-3233.7638
x8	-1862.0178
x9	-483.4178
x10	-1309.4675
x11	3625.5745
x12	4731.6672
x13	-4312.8244
x14	-3060.1174
x15	-2051.3965
x16	-5435.812
x17	811.334
x18	4813.9169
x19	6590.7194

Adj. R-squared: 0.918



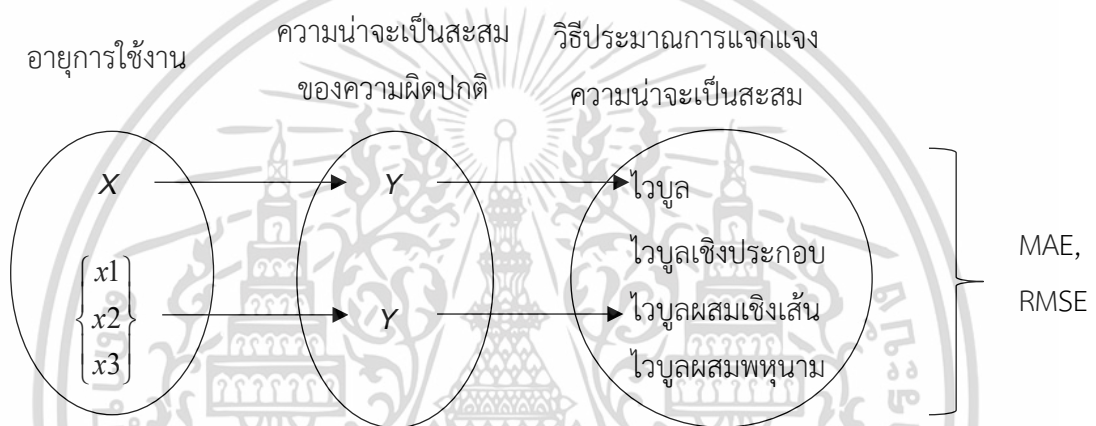
Jarque Bera: (statistic=0.5049748115896134, p value=0.776865996787474)

รูปที่ 4.5 แผนภูมิแสดงความผันแปรของส่วนเหลือที่ได้จากการแจกแจงไวบูลผสมพหุนามอันดับ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลองการแจกแจงไวบูลผสมพหุนามอันดับ 2 ในตารางที่ 4.7 และพหุนามอันดับ 3 ในตารางที่ 4.8 พบว่าค่าสัมประสิทธิ์การกำหนดพหุคูณที่ปรับแก้แล้ว (Adjusted R Square) มีค่าเท่ากันคือ 0.918 และการแจกแจงของส่วนเหลือจากสถิติทดสอบจาร์คูปรีามีค่าพีไม่แตกต่างกันดังรูปที่ 4.4 และ 4.5 ผู้วิจัยจึงจะเลือกการแจกแจงไวบูลผสมพหุนามอันดับ 2 ในการประมาณค่าความน่าจะเป็นสะสมความผิดปกติของเครื่องจักรเพราะมีความซับซ้อนของตัวแบบน้อยกว่า

4.1.3 ทำการเปรียบเทียบประสิทธิภาพของการแจกแจงสะสมทั้ง 4 วิธี ด้วยวิธีความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย เพื่อหาการแจกแจงสะสมที่เหมาะสมโดยจะพิจารณาจากค่าที่ต่ำสุด สามารถสรุปได้ตามรูปที่ 4.6



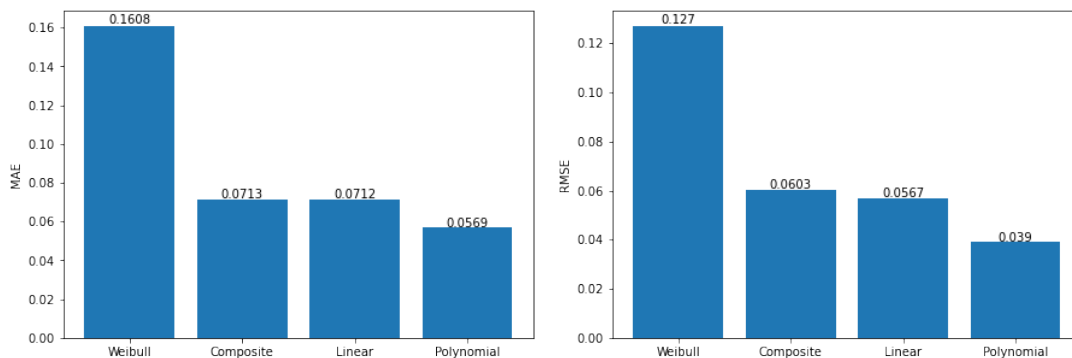
รูปที่ 4.6 แผนภูมิแสดงความสัมพันธ์ของตัวแบบที่ได้จากข้อมูลการบำรุงรักษา

จากรูปที่ 4.6 จะพบว่าวิธีไวบูลพิจารณาความน่าจะเป็นสะสมของความผิดปกติจากข้อมูลการบำรุงรักษาทั้งหมด ในขณะที่วิธีไวบูลเชิงประกอบ ไวบูลผสมเชิงเส้น และไวบูลผสมพหุนามพิจารณาความน่าจะเป็นสะสมของความผิดปกติจากข้อมูลการบำรุงรักษาที่แยกตามรูปแบบความผิดปกติ

ตารางที่ 4.9 ความน่าจะเป็นเป็นอิสระของความผิดพลาดแต่ละวิธี

อายุการใช้งาน	ความน่าจะเป็นอิสระของความผิดพลาด	ความน่าจะเป็นอิสระของความผิดพลาดที่ทำนายได้			
		วิธีการแจกแจงไวบูล	วิธีการแจกแจงไวบูลเชิงประกอบ	วิธีการแจกแจงไวบูลผสมเชิงเส้น	วิธีการแจกแจงไวบูลผสมพหุนาม
61	0.045455	0.02629	0.000037	0.056309	0.045455
273	0.11039	0.377751	0.197457	0.190266	0.142857
273	0.175325	0.377751	0.197457	0.190266	0.142857
276	0.24026	0.383984	0.208927	0.203457	0.272727
276	0.305195	0.383984	0.208927	0.203457	0.272727
304	0.37013	0.441963	0.336668	0.349666	0.37013
334	0.435065	0.502898	0.507633	0.540818	0.467532
334	0.5	0.502898	0.507633	0.540818	0.467532
335	0.564935	0.504896	0.513697	0.547479	0.564935
365	0.62987	0.563486	0.736744	0.74177	0.75974
365	0.694805	0.563486	0.736744	0.74177	0.75974
365	0.75974	0.563486	0.736744	0.74177	0.75974
365	0.824675	0.563486	0.736744	0.74177	0.75974
365	0.88961	0.563486	0.736744	0.74177	0.75974
699	0.954545	0.944367	1	0.968617	0.954545

จากตารางที่ 4.9 จะพบว่าทั้ง 4 วิธีสามารถทำนายความน่าจะเป็นอิสระของความผิดพลาดได้สอดคล้องกับความน่าจะเป็นอิสระของความผิดพลาดจากวิธีของปีนาร์ด และมีค่าความน่าจะเป็นอิสระอยู่ระหว่าง 0 และ 1 ตามสมบัติตามความน่าจะเป็นอิสระ



รูปที่ 4.7 แผนภูมิแสดงค่าเฉลี่ยของความผิดพลาดสัมบูรณ์ และค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของการแจกแจงแต่ละวิธี

จากรูปที่ 4.7 วิธีการแจกแจงไวบูลผสมพหุนามมีความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดเท่ากับ 0.056 และ 0.039 ตามลำดับ กล่าวคือมีประสิทธิภาพในการทำนายความน่าจะเป็นสะสมของความผิดปกติสูงที่สุด

4.2 ผลการวิจัยข้อมูลการใช้งานเครื่องจักร

4.2.1 การเตรียมและตรวจสอบข้อมูล

ในงานวิจัยนี้ได้พัฒนาตัวแบบจากข้อมูลการใช้งานเครื่องจักรแบ่งเป็น 2 ประเภทคือ ข้อมูลที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร และข้อมูลที่เกี่ยวข้องกับความเร็วรอบของเครื่องจักร ซึ่งการแบ่งประเภททั้ง 2 ประเภท จะพิจารณาจากค่าสหสัมพันธ์ระหว่างข้อมูลกับความเร็วรอบของเครื่องจักร (correlation, R) และจำนวนข้อมูล (n) ได้ผลดังตาราง

ตารางที่ 4.10 จำนวนตัวแบบการใช้งานเครื่องจักรในแต่ละประเภทของตัวแบบ

ประเภทของตัวแบบ	จำนวนตัวแบบทั้งหมด
ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร	30
สัมพันธ์กับความเร็วรอบของเครื่องจักร	70

จากตารางที่ 4.10 แสดงจำนวนตัวแบบที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักรจำนวน 30 ตัวแบบและตัวแบบที่สัมพันธ์กับความเร็วรอบของเครื่องจักรจำนวน 70 ตัวแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 การพัฒนาตัวแบบไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร

ข้อมูลในตัวแบบที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร นี้จะถูกแปลงข้อมูลด้วยวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ ให้อยู่ในรูปแบบการแจกแจงปรกติก่อน จากการประมาณค่าสัมประสิทธิ์ของบ็อกซ์-ค็อกซ์ และตรวจสอบการแจกแจงปรกติด้วยวิธีจาร์คูบีรา

H0: ตัวแบบการใช้งานเครื่องจักรมีการแจกแจงปรกติ

H1: ตัวแบบการใช้งานเครื่องจักรไม่มีการแจกแจงปรกติ

ที่ความเชื่อมั่นร้อยละ 95 จะได้ผลลัพธ์ดังตารางที่ 4.11

ตารางที่ 4.11 ตัวแบบการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร

ตัวแบบ	ค่าพีจากการทดสอบจาร์คูบีราก่อนแปลง	ค่าสัมประสิทธิ์ของบ็อกซ์-ค็อกซ์	ค่าพีจากการทดสอบจาร์คูบีราหลังแปลง
TSO-ENGIN_A-4200-TV -351X -DIRECT	0.0000	0.793654	0.0000
TSO-ENGIN_A-4200-TV -352X -DIRECT	0.5700	0.218233	0.9800
TSO-ENGIN_A-4200-TV -353Y -DIRECT	0.0000	-3.427713	0.0400
TSO-LUBE_A -4200-TE -878A -PV	0.0000	0.419358	0.0000
TSO-LUBE_A -4200-TE -900A -PV	0.0000	1.527228	0.0000
TSO-ENGIN_A-4200-TV -354X -DIRECT	0.0000	-1.523179	0.0000
TSO-ENGIN_A-4200-TV -354Y -DIRECT	0.0000	-4.543849	0.4500
TSO-ENGIN_A-4200-TV -355X -DIRECT	0.0000	-5.296061	0.9000
TSO-DGS_A -4200-PDT-726A -PV	0.0000	-0.540856	0.8700
TSO-DGS_A -4200-PDT-727A -PV	0.0000	-0.119494	0.0000
TSO-DGS_A -4200-PT -720A -PV	0.2700	-1.126759	0.9800
TSO-DGS_A -4200-PT -737A -PV	0.5700	1.01571	0.5700
TSO-ENCAN_A-4200-TE -903A -PV	0.0000	-2.769376	0.9300
AP_HPC_ETA_Act	0.1700	1.663314	0.3400
AP_HPC_ETA_Nom	0.0000	-30.985647	0.0000
AP_HPC_Spd_Nom	0.0000	-19.700843	0.0000
TSO-COMU_A -4200-XT -752A -DIRECT	0.0000	-8.817765	0.0200
TSO-COMU_A -4200-YT -753A -DIRECT	0.0000	-7.812008	0.9900
TSO-FUEL_A -4200-PT -900A -PV	0.0000	84.222065	0.0000
TSO-FUEL_A -4200-TT -900A -PV	0.0000	-3.375578	0.2100
TSO-STATN -4200-FV -411 -PV	0.0000	713.934939	0.0000
TSO-YARD_A -4200-RT -563 -PV	0.0000	30.448138	0.0000
TSO-FUEL_A -4200-PDT-896A -PV	0.0000	0.008879	0.0100
TSO-FUEL_A -4200-PT -892A -PV	0.0000	8.426742	0.0100
TSO-FUEL_A -4200-PT -895A -PV	0.0000	54.748164	0.0000
TSO-FUEL_A -4200-PT -905A -PV	0.0000	58.616535	0.0000
TSO-FUEL_A -4200-PT -936A -PV	0.0200	0.595515	0.0200
TSO-FUEL_A -4200-TE -902A -PV	0.0000	-19.940824	0.0000

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.11 แสดงให้เห็นค่าพีที่ได้จากการทดสอบจาร์คูปีราของแต่ละตัวแบบ ซึ่งจะเห็นว่าตัวแบบที่มีการแจกแจงปกติมีจำนวน 10 ตัวแบบที่สามารถนำไปพัฒนาตัวแบบต่อได้ ในงานวิจัยนี้จะพัฒนาตัวแบบด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและวิธีป่าไม้โดดเดี่ยว

1. วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว (Univariate k-sigma)

วิธีนี้จะกำหนดขีดจำกัดควบคุมได้จากค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของแต่ละตัวแบบดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \bar{Y} + 3 S.D.$$

$$\text{เส้นกึ่งกลาง} = \text{ค่าเฉลี่ย } \bar{Y}$$

$$\text{ขีดจำกัดควบคุมล่าง} = \bar{Y} - 3 S.D$$

$$S.D. = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}}$$

ตารางที่ 4.12 ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และขีดจำกัดควบคุมของตัวแบบการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร

ตัวแบบ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ขีดจำกัดควบคุม
TSO-ENGIN_A-4200-TV -352X -DIRECT	-0.81	0.04	[-0.92-0.7]
TSO-ENGIN_A-4200-TV -354Y -DIRECT	-61.94	6.62	[-81.81-42.06]
TSO-ENGIN_A-4200-TV -355X -DIRECT	-10.55	2.94	[-19.37-1.73]
TSO-DGS_A -4200-PDT-726A -PV	-2.74	0.22	[-3.4-2.08]
TSO-DGS_A -4200-PT -720A -PV	0.89	0.00	[0.89.0.89]
TSO-DGS_A -4200-PT -737A -PV	-0.36	0.05	[-0.53-0.2]
TSO-ENCAN_A-4200-TE -903A -PV	0.36	0.00	[0.36.0.36]
AP_HPC_ETA_Act	-0.46	0.02	[-0.51-0.42]
TSO-COMU_A -4200-YT -753A -DIRECT	-27.96	5.55	[-44.61-11.32]
TSO-FUEL_A -4200-TT -900A -PV	0.30	0.00	[0.3.0.3]

จากตารางที่ 4.12 จะพบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว ตัวแบบจะมีขีดจำกัดควบคุมที่ยอมรับได้อยู่ในช่วงค่าเฉลี่ย ± 3 เท่าของส่วนเบี่ยงเบนมาตรฐาน

2. วิธีป่าไม้โดดเดี่ยว (Isolation Forest)

ในงานวิจัยนี้กำหนดค่าไฮเพอร์พารามิเตอร์ของตัวแบบดังนี้

ค่าความไม่บริสุทธิ์ของข้อมูล (contamination) เป็นแบบอัตโนมัติ ซึ่งตัวแบบจะพิจารณาค่าความน่าจะเป็นของความไม่บริสุทธิ์ของข้อมูล ตามความน่าจะเป็นของข้อมูลที่มีค่าสูงหรือต่ำกว่า

ข้อมูลส่วนใหญ่ (Outlier)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

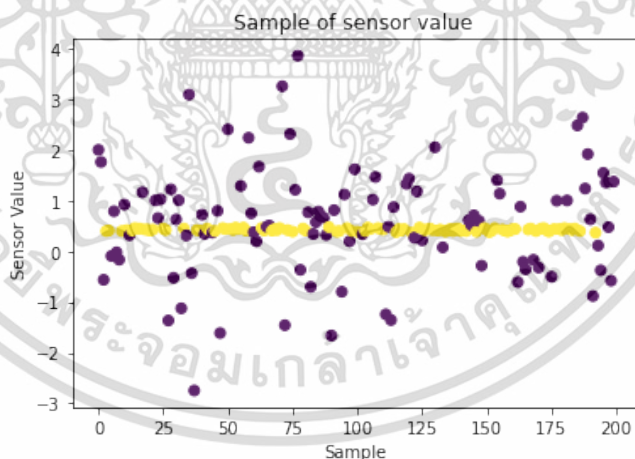
จำนวนตัวประมาณค่าของตัวแบบ (estimator) กล่าวคือจำนวนต้นไม้ตัดสินใจในตัวแบบป่าไม้โอดเดี๋ยว เป็น 10 ตัวประมาณค่า

```
{'bootstrap': False,
'contamination': 'auto',
'max_features': 1.0,
'max_samples': 'auto',
'n_estimators': 10,
'n_jobs': None,
'random_state': 0,
'verbose': 0,
'warm_start': False}
```

รูปที่ 4.8 ข้อมูลแสดงการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบวิธีป่าไม้โอดเดี๋ยว

จากรูปที่ 4.8 จะพบว่าค่าไฮเปอร์พารามิเตอร์ที่สำคัญของตัวแบบจะตั้งตามค่ามาตรฐาน ยกเว้นจำนวนตัวประมาณค่าของตัวแบบ จะขึ้นอยู่กับความซับซ้อนของตัวแบบและจำเป็นต้องมีการทดลองและปรับปรุงให้ตัวแบบมีประสิทธิภาพสูงสุด

การทดสอบประสิทธิภาพของตัวแบบวิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและตัวแบบวิธีป่าไม้โอดเดี๋ยว ด้วยการสุ่มข้อมูลการใช้งานในสภาวะปกติ 100 ตัวอย่างและสังเคราะห์ข้อมูลการใช้งานในสภาวะไม่ปกติ 100 ตัวอย่าง รวม 200 ตัวอย่าง



รูปที่ 4.9 ข้อมูลทดสอบการใช้งานในสภาวะปกติและผิดปกติ

จากรูปที่ 4.9 แสดงให้เห็นข้อมูลการใช้งานเครื่องจักรที่สภาวะการใช้งานปกติและผิดปกติ โดยจุดสีเหลืองจะแสดงข้อมูลในสภาวะการใช้งานปกติและจุดสีม่วงแสดงข้อมูลในสภาวะการใช้งานผิดปกติ เมื่อนำข้อมูลทั้ง 200 ตัวอย่าง มาทดสอบจะได้ผลการทดสอบดังตารางที่ 4.13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

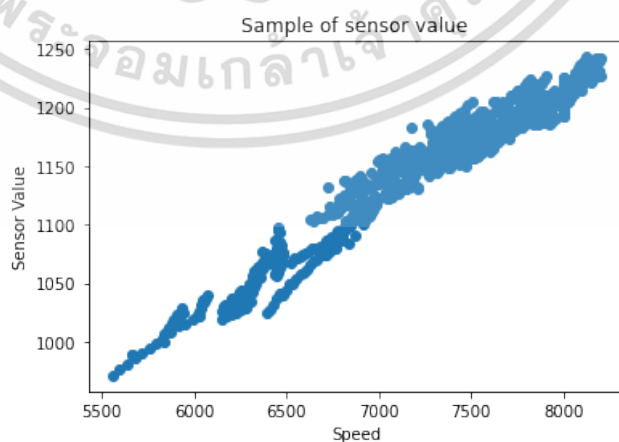
ตารางที่ 4.13 แสดงผลการเปรียบเทียบค่าความเที่ยง ค่าเรียกคืน และคะแนน F1

ตัวแบบ	ตัวแบบวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียว			ตัวแบบวิธีป่าไม้โตดเดี่ยว		
	ค่าความเที่ยง	ค่าเรียกคืน	คะแนน F1	ค่าความเที่ยง	ค่าเรียกคืน	คะแนน F1
TSO-ENGIN_A-4200-TV -352X -DIRECT	0.93	1.00	0.97	0.95	0.86	0.90
TSO-ENGIN_A-4200-TV -354Y -DIRECT	0.50	0.99	0.66	0.64	0.97	0.77
TSO-ENGIN_A-4200-TV -355X -DIRECT	0.50	1.00	0.67	0.60	0.85	0.71
TSO-DGS_A -4200-PDT-726A -PV	0.67	0.99	0.80	0.73	0.95	0.82
TSO-DGS_A -4200-PT -720A -PV	1.00	1.00	1.00	1.00	0.85	0.92
TSO-DGS_A -4200-PT -737A -PV	0.94	1.00	0.97	0.94	0.95	0.95
TSO-ENCAN_A-4200-TE -903A -PV	1.00	1.00	1.00	1.00	0.95	0.97
AP_HPC_ETA_Act	0.95	1.00	0.98	0.98	0.88	0.93
TSO-COMU_A -4200-YT -753A -DIRECT	0.50	0.99	0.66	0.77	0.92	0.84
TSO-FUEL_A -4200-TT -900A -PV	1.00	1.00	1.00	1.00	0.87	0.93
Average	0.80	1.00	0.87	0.86	0.91	0.87

จากตารางที่ 4.13 จะเห็นได้ว่าค่าเฉลี่ยของตัวแบบวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวมีค่าเรียกคืนเท่ากับ 1.00 และค่าคะแนน F1 เท่ากับ 0.87 สูงกว่าตัวแบบตัวแบบวิธีป่าไม้โตดเดี่ยว ในงานวิจัยนี้จะใช้วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวในการทำนายความผิดปกติของตัวแบบการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักร

4.2.3 การพัฒนาตัวแบบตัวแบบสัมพันธ์กับความเร็วรอบของเครื่องจักร

ข้อมูลของตัวแบบสัมพันธ์กับความเร็วรอบของเครื่องจักรจะถูกนำมาตรวจสอบตรวจสอบสมมติฐานของสมการถดถอยเชิงเส้นก่อน ซึ่งในงานวิจัยนี้จะพิจารณาเฉพาะตัวแบบที่ผ่านตามสมมติฐานเท่านั้น



รูปที่ 4.10 ข้อมูลการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.10 แสดงตัวอย่างข้อมูลของข้อมูลของตัวแบบสัมพันธ์กับความเร็วรอบของเครื่องจักร พบว่าสัญญาณอุณหภูมิของเครื่องจักร (องศาฟาเรนไฮต์) ที่แปรผันตามความเร็วรอบของเครื่องจักร (rpm)

ตารางที่ 4.14 ผลการตรวจสอบสมมุติฐานของสมการถดถอยเชิงเส้นระหว่างข้อมูลการใช้งานเครื่องจักรกับความเร็วยรอบเครื่องจักร

ตัวแบบ	ค่าสัมประสิทธิ์สหสัมพันธ์	ค่าทดสอบเดออร์บินวัตสัน	ค่าพีจากการการทดสอบไวท์	ค่าพีจากการทดสอบจาร์คูเปร์รา
TSO-DGS_A -4200-PT -736A -PV	0.2155	0.0381	0.0000	0.0021
TSO-ENGIN_A-4200-ST -438 -PV	0.9401	0.0268	0.0000	0.0000
TSO-ENGIN_A-4200-TE -851A -PV	0.8659	0.0571	0.0000	0.0001
TSO-ENGIN_A-4200-TE -852A -PV	0.8908	0.0836	0.0000	0.0001
TSO-ENGIN_A-4200-TE -853A -PV	0.8726	0.0607	0.0000	0.0019
TSO-ENGIN_A-4200-TE -854A -PV	0.8043	1.5666	0.0610	0.1978
TSO-ENGIN_A-4200-TE -855A -PV	0.8940	0.0536	0.0000	0.0001
TSO-ENGIN_A-4200-TE -856A -PV	0.7666	0.1008	0.0000	0.0000
TSO-ENGIN_A-4200-TE -857A -PV	0.8635	0.0642	0.0000	0.0068
TSO-ENGIN_A-4200-TE -858A -PV	0.8613	0.0513	0.0005	0.0175
TSO-ENGIN_A-4200-TE -859A -PV	0.8777	0.0650	0.0000	0.0001
TSO-ENGIN_A-4200-TE -860A -PV	0.8757	0.0737	0.0000	0.0001
TSO-ENGIN_A-4200-TE -861A -PV	0.8836	0.0713	0.0000	0.0004
TSO-ENGIN_A-4200-TE -862A -PV	0.8736	0.0548	0.0000	0.0005
TSO-ENGIN_A-4200-TE -863A -PV	0.8734	0.0357	0.0840	0.0001
TSO-ENGIN_A-4200-TE -864A -PV	0.8705	0.0772	0.0000	0.0000
TSO-ENGIN_A-4200-TE -865A -PV	0.8789	0.0659	0.0000	0.0000
TSO-ENGIN_A-4200-TE -866A -PV	0.8719	0.0588	0.0001	0.0002
TSO-ENGIN_A-4200-TE -867A -PV	0.7368	0.1208	0.0009	0.0000
TSO-ENGIN_A-4200-TE -949A -PV	0.7756	0.0412	0.0000	0.0000
TSO-ENGIN_A-4200-TV -351Y -DIRECT	0.6077	0.0719	0.0540	0.0998
TSO-ENGIN_A-4200-TV -351Z -DIRECT	0.4845	0.1134	0.0000	0.0002
TSO-ENGIN_A-4200-TV -352Y -DIRECT	0.8356	0.1949	0.0000	0.0088
TSO-ENGIN_A-4200-TV -353X -DIRECT	0.2391	0.1687	0.0000	0.0000
TSO-ENGIN_A -4200-EFF-THERMAL	0.7950	1.6177	0.0000	0.0000
TSO-FUEL_A -4200-FCV-964A -FB	0.9597	1.8312	0.0840	0.1017
TSO-LUBE_A -4200-TE -871A -PV	0.3862	0.0677	0.0000	0.0000
TSO-LUBE_A -4200-TE -875A -PV	0.1417	0.1536	0.0000	0.0000
TSO-LUBE_A -4200-TE -876A -PV	0.4315	0.0572	0.0000	0.0000
UAVFD430_P750_P011_DCBusVoltage	0.1631	0.1127	0.0000	0.0000
TSO-ENGIN_A-4200-TV -350_1-DIRECT	0.7969	0.2114	0.0000	0.0000
TSO-LUBE_A -4200-PDT-908A -PV	0.2556	1.8312	0.0000	0.0000
TSO-LUBE_A -4200-PT -901A -PV	0.6788	0.0990	0.0003	0.0000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.14 ผลการตรวจสอบสมมติฐานของสมการถดถอยเชิงเส้นระหว่างข้อมูลการใช้งาน
เครื่องจักรเทียบกับความเร็วรอบเครื่องจักร (ต่อ)

ตัวแบบ	ค่าสัมประสิทธิ์ สหสัมพันธ์	ค่าทดสอบ เดอ์บินวัตสัน	ค่าพิจากการการ ทดสอบไวท์	ค่าพิจากการ ทดสอบจาร์คูปรี่า
TSO-LUBE_A -4200-PT -909A -PV	0.1256	0.0979	0.0000	0.0000
TSO-LUBE_A -4200-PT -908A -PV	0.4314	0.0286	0.0275	0.0000
TSO-LUBE_A -4200-TE -877A -PV	0.2209	0.0575	0.0000	0.0000
TSO-LUBE_A -4200-TE -907A -PV	0.2528	0.0605	0.0000	0.0000
TSO-ENGIN_A-4200-ST -439 -PV	0.9951	0.3429	0.7328	0.0103
TSO-ENGIN_A-4200-TV -355Y -DIRECT	0.5553	0.2373	0.0000	0.0000
TSO-ENGIN_A-4200-TV -355Z -GAP	0.3943	0.0967	0.0000	0.0000
TSO-LUBE_A -4200-TE -870A -PV	0.3930	0.0802	0.0000	0.0000
TSO-LUBE_A -4200-TE -879A -PV	0.3411	0.0710	0.0000	0.0000
TSO-DGS_A -4200-PDT-723A -PV	0.2380	0.8611	0.0000	0.0004
TSO-DGS_A -4200-PT -731A -PV	0.2175	0.0683	0.0000	0.0002
TSO-DGS_A -4200-PY -722A -PV	0.5917	0.1489	0.0000	0.0000
TSO-ENGIN_A-4200-PDT-799 -PV	0.8278	1.6643	0.0640	0.1184
AP_HPC_Flow_Act	0.8942	0.1562	0.0000	0.0000
AP_HPC_Head	0.5269	0.0672	0.0000	0.4928
TSO-COMU_A -4200-TE -741A -PV	0.1823	0.0634	0.0000	0.0000
TSO-COMU_A -4200-TE -742A -PV	0.4763	0.0479	0.0000	0.0000
TSO-COMU_A -4200-TE -743A -PV	0.4758	0.0504	0.0000	0.0000
TSO-COMU_A -4200-TE -744A -PV	0.3839	0.0474	0.0000	0.0000
TSO-COMU_A -4200-XT -750A -DIRECT	0.5489	0.4971	0.0000	0.0000
TSO-COMU_A -4200-YT -751A -DIRECT	0.6051	0.6557	0.0000	0.2109
TSO-FUEL_A -4200-FT -900A -PV	0.7715	0.0914	0.0000	0.0001
TSO-YARD_A -4200-PDT-425 -PV	0.6290	0.0870	0.0000	0.0000
AP_Eng_Fuel_Flow_Std_GF	0.9342	1.6333	0.0520	0.6959
AP_Eng_PCD_Pred_GF	0.9522	0.2666	0.0000	0.0001
AP_Eng_Pwr_Pred_GF	0.9570	0.1254	0.0000	0.0005
AP_Eng_T5_Pred_GF	0.8979	0.1299	0.0000	0.0000
TSO-FUEL_A -4200-PDT-586A -PV	0.7757	0.0915	0.0000	0.0001
TSO-FUEL_A -4200-FCV-964A -CMD	0.9574	1.6022	0.0720	0.1783
TSO-ENCAN_A-4200-R201_R202	0.4427	0.0240	0.0204	0.0000
TSO-ENGIN_A-4200-EBV-942A -FB	0.1393	0.0086	0.0000	0.0000
TSO-ENGIN_A-4200-EGV-943A -FB	0.9245	0.0351	0.0000	0.0000
TSO-ENGIN_A-4200-MC -943A -PV	0.4554	0.7015	0.0000	0.0044
TSO-ENGIN_A-4200-EGV-943A -CMD	0.9626	0.0244	0.0000	0.0000
TSO-ENGIN_A-4200-OSM_BU	0.9929	0.3513	0.1855	0.0001
TSO-ENGIN_A-4200-PT -903A -PV	0.9750	0.0439	0.0000	0.0002

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.14 แสดงให้เห็นค่าสัมประสิทธิ์พหุคูณ และค่าพีที่ได้จากการทดสอบจาร์คูบราของส่วนเหลือในแต่ละตัวแบบ ซึ่งจะเห็นว่าทุกตัวแบบมีความสัมพันธ์เชิงเส้นกับความเร็วรอบเครื่องจักร แต่ส่วนเหลือที่มีการแจกแจงปรกติมีเพียง 5 ตัวแบบ ที่จะใช้ในการพัฒนาตัวแบบต่อไปได้ ในงานวิจัยนี้จะพัฒนาตัวแบบด้วยวิธีแผนภูมิควบคุมถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธีที่เสนอคือวิธี k เท่าของส่วนเหลือมาตรฐาน (k standard residual)

1) วิธีแผนภูมิควบคุมการถดถอย (Regression control chart)

วิธีนี้จะกำหนดขีดจำกัดควบคุมที่ได้จากค่าคลาดเคลื่อนมาตรฐาน (Standard error) และค่าสัมประสิทธิ์การกำหนดพหุคูณ (R Square) ในแต่ละตัวแบบดังนี้

$$\text{ขีดจำกัดควบคุมบน} = \hat{Y}_i + 3 S_{error}$$

$$\text{เส้นกึ่งกลาง} = \text{ค่าทำนาย } \hat{Y}_i$$

$$\text{ขีดจำกัดควบคุมล่าง} = \hat{Y}_i - 3 S_{error}$$

$$S_{error} = S.D. \cdot \sqrt{1-r^2}$$

$$S.D. = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}}$$

ตารางที่ 4.15 ขีดจำกัดควบคุมของตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักรวิธีแผนภูมิควบคุมถดถอย

ตัวแบบ	ขีดจำกัดควบคุมล่าง	เส้นกึ่งกลาง	ขีดจำกัดควบคุมบน
TSO-ENGIN_A-4200-TE-854A-PV	0.09 x ความเร็ว + 552.57	0.09 x ความเร็ว + 553.28	0.09 x ความเร็ว + 553.98
TSO-FUEL_A-4200-FCV-964A-FB	0.01 x ความเร็ว - 4.81	0.01 x ความเร็ว - 4.78	0.01 x ความเร็ว - 4.75
TSO-ENGIN_A-4200-PDT-799-PV	0 x ความเร็ว - 0.72	0 x ความเร็ว - 0.72	0 x ความเร็ว - 0.71
AP_Eng_Fuel_Flow_Std_GF	0.02 x ความเร็ว - 91.23	0.02 x ความเร็ว - 90.84	0.02 x ความเร็ว - 90.46
TSO-FUEL_A-4200-FCV-964A-CMD	0.01 x ความเร็ว - 5.12	0.01 x ความเร็ว - 5.09	0.01 x ความเร็ว - 5.06

จากตารางที่ 4.15 จะพบว่าวิธีแผนภูมิควบคุมถดถอยตัวแบบจะมีขีดจำกัดควบคุมที่ยอมรับได้เท่ากับค่าทำนาย ± 3 เท่าของค่าคลาดเคลื่อนมาตรฐาน

2) วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร (Multi-variate k-sigma)

วิธีนี้สามารถคำนวณได้จากระยะทางมาฮาလာโนบิส โดยพิจารณาค่า k_i หากมีค่ามากขึ้นจะมีโอกาสที่ข้อมูลผิดปกติมากขึ้นด้วย

$$k_i = \sqrt{(\bar{X}_i - \bar{\pi})^T \Sigma^{-1} (\bar{X}_i - \bar{\pi})}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 ค่าเฉลี่ยและค่าความแปรปรวนร่วมผกผันกับความเร็รรอบของเครื่องจักร

ตัวแบบ	ค่าเฉลี่ย ($\bar{\pi}$)	ค่าความแปรปรวนร่วมผกผัน (Σ^{-1})
TSO-ENGIN_A-4200-TE -854A -PV	1219.975261	[[0.007081978153078225, -0.0006032052860359252...
TSO-FUEL_A -4200-FCV-964A -FB	48.345545	[[7.102130784597069, -0.04820277374665596], [-...
TSO-ENGIN_A-4200-PDT-799 -PV	0.963081	[[1327.946922508751, -0.2848170074383573], [-0...
AP_Eng_Fuel_Flow_Std_GF	103.393081	[[0.3126417644478856, -0.00775820796407292], [...
TSO-FUEL_A -4200-FCV-964A -CMD	48.039542	[[6.698907687397641, -0.04546841133385102], [-...

จากตารางที่ 4.16 แสดงค่าเฉลี่ยและค่าความแปรปรวนร่วมผกผันกับความเร็รรอบของเครื่องจักรที่ใช้ในการคำนวณระยะทางมาฮาလာโนบิส

3) วิธี k เท่าของส่วนเหลือมาตรฐาน (k standard residual)

วิธีนี้จะกำหนดขีดจำกัดควบคุมจากการแจกแจงของส่วนเหลือและค่าทำนายของตัวแบบ

$$\text{ขีดจำกัดควบคุมบน} = \hat{Y}_i + 3 S_{residual}$$

$$\text{เส้นกึ่งกลาง} = \hat{Y}_i$$

$$\text{ขีดจำกัดควบคุมล่าง} = \hat{Y}_i - 3 S_{residual}$$

$$S_{residual} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}}$$

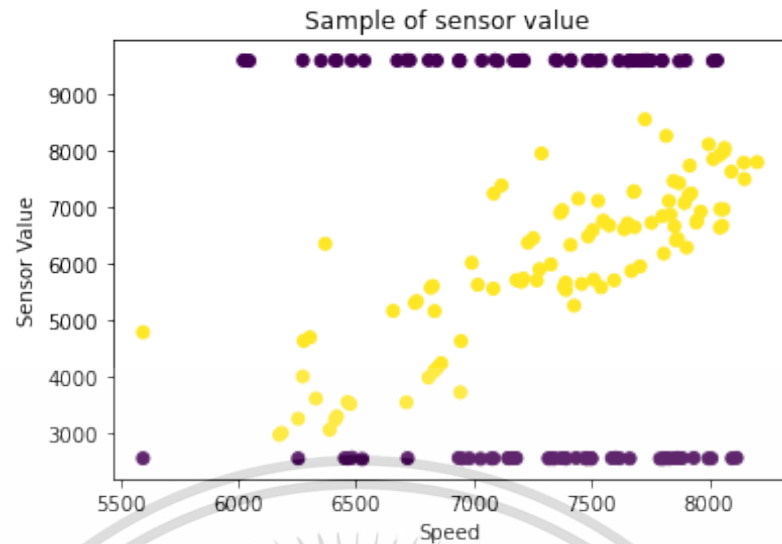
ตารางที่ 4.17 ขีดจำกัดควบคุมของตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็รรอบของเครื่องจักร วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐาน

ตัวแบบ	ขีดจำกัดควบคุมล่าง	เส้นกึ่งกลาง	ขีดจำกัดควบคุมบน
TSO-ENGIN_A-4200-TE -854A -PV	0.09 x ความเร็ว + 517.6	0.09 x ความเร็ว + 553.28	0.09 x ความเร็ว + 588.95
TSO-FUEL_A -4200-FCV-964A -FB	0.01 x ความเร็ว - 5.91	0.01 x ความเร็ว - 4.78	0.01 x ความเร็ว - 3.65
TSO-ENGIN_A-4200-PDT-799 -PV	0 x ความเร็ว - 0.8	0 x ความเร็ว - 0.72	0 x ความเร็ว - 0.63
AP_Eng_Fuel_Flow_Std_GF	0.02 x ความเร็ว - 96.21	0.02 x ความเร็ว - 90.84	0.02 x ความเร็ว - 85.47
TSO-FUEL_A -4200-FCV-964A -CMD	0.01 x ความเร็ว - 6.25	0.01 x ความเร็ว - 5.09	0.01 x ความเร็ว - 3.93

จากตารางที่ 4.17 จะพบว่าวิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐานจะมีขีดจำกัดควบคุมที่ยอมรับได้เท่ากับค่าทำนาย ± 3 เท่าของการแจกแจงของส่วนเหลือ

การทดสอบประสิทธิภาพของตัวแบบวิธีแผนภูมิควบคุมถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐาน ด้วยการสุ่มข้อมูลการใช้งานในสภาวะปกติ 100 ตัวอย่างและสังเคราะห์ข้อมูลการใช้งานในสภาวะไม่ปกติ 100 ตัวอย่าง รวมทั้ง 200 ตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 ข้อมูลทดสอบการใช้งานในสภาวะปกติและผิดปกติ

จากรูปภาพที่ 4.11 แสดงให้เห็นข้อมูลที่สภาวะการใช้งานปกติและผิดปกติ โดยแบ่งเป็นจุดสีเหลืองแสดงข้อมูลในสภาวะการใช้งานปกติ และจุดสีม่วงแสดงข้อมูลในสภาวะการใช้งานผิดปกติ เมื่อนำข้อมูลทั้ง 200 ตัวอย่าง มาทดสอบจะได้ผลการทดสอบดังตารางที่ 4.18

ตารางที่ 4.18 ผลการเปรียบเทียบค่าความเที่ยง ค่าเรียกคืน และคะแนน F1

ตัวแบบ	ตัวแบบวิธีแผนภูมิควบคุมถดถอย			วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร			วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐาน		
	ค่าความเที่ยง	ค่าเรียกคืน	คะแนน F1	ค่าความเที่ยง	ค่าเรียกคืน	คะแนน F1	ค่าความเที่ยง	ค่าเรียกคืน	คะแนน F1
TSO-ENGIN_A-4200-TE-854A -PV	1.00	0.15	0.26	0.53	0.52	0.53	1.00	1.00	1.00
TSO-FUEL_A -4200-FCV-964A -FB	1.00	0.28	0.44	0.49	0.52	0.50	0.93	1.00	0.96
TSO-ENGIN_A-4200-PDT-799 -PV	1.00	0.51	0.68	0.51	0.54	0.53	0.98	1.00	0.99
AP_Eng_Fuel_Flow_Std_GF	1.00	0.21	0.35	0.54	0.58	0.56	0.96	1.00	0.98
TSO-FUEL_A -4200-FCV-964A -CMD	0.98	0.39	0.56	0.54	0.60	0.57	0.96	1.00	0.98
Average	1.00	0.31	0.46	0.52	0.55	0.54	0.97	1.00	0.98

จากตารางที่ 4.18 จะเห็นได้ว่าค่าเฉลี่ยตัวแบบวิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐาน โดยรวมมีค่าความเที่ยง ค่าเรียกคืน และคะแนน F1 สูงกว่าตัวแบบวิธีขีดจำกัดควบคุมด้วยวิธีแผนภูมิควบคุมการถดถอยและตัวแบบวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร ในงานวิจัยนี้จะใช้วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐานในการทำนายความ

ผิดปกติของตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักร เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การหาค่าความผิดปกติร่วม

4.3.1 การคำนวณค่าทำนายของตัวแบบในรูปแบบความน่าจะเป็น

จากการทดลองเปรียบเทียบประสิทธิภาพของตัวแบบการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักรและตัวแบบการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็วรอบของเครื่องจักรจะได้ว่าวิธีที่มีประสิทธิภาพดีที่สุดคือ วิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและวิธีขีดจำกัดควบคุมด้วยวิธี k เท่าของส่วนเหลือมาตรฐานตามลำดับ

ในงานวิจัยนี้จะพิจารณาค่าความน่าจะเป็นจากการแจกแจงปกติของค่าคะแนนมาตรฐาน

$$P_{operation} = \begin{cases} s(k_i - c), k_i = |z| \geq c \\ 0, otherwise \end{cases}$$

ตารางที่ 4.19 ค่าความผิดปกติงานข้อมูลการใช้งานเครื่องจักรแต่ละตัวแบบ

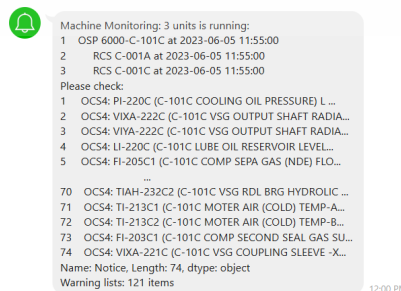
เวลบบันทึก	เลขที่ตัวแบบ	ค่าความผิดปกติ
2023-03-26-16.00.00.000000	167	0
2023-03-26-16.00.00.000000	168	94
2023-03-26-16.00.00.000000	169	7
2023-03-26-16.00.00.000000	170	6
2023-03-26-16.00.00.000000	171	89
2023-03-26-16.00.00.000000	172	0
2023-03-26-16.00.00.000000	173	100

จากตารางที่ 4.19 คือตัวอย่างการแสดงค่าความผิดปกติของแต่ละตัวแบบที่ได้จากการคำนวณ โดยเมื่อนำไปใช้งานแต่ละตัวแบบจะคำนวณค่าความผิดปกติของจุดข้อมูลใหม่ตามรอบที่กำหนด เช่น คำนวณค่าความผิดปกติทุก 5 นาที เป็นต้น

4.3.2 การคำนวณความน่าจะเป็นของความผิดปกติร่วม

เนื่องจากข้อมูลการบำรุงรักษาจากการแจกแจงไวบูลผสมพหุนามและข้อมูลการใช้งานเครื่องจักรเป็นอิสระต่อกัน จึงสามารถหาความน่าจะเป็นของความผิดปกติร่วมได้ดังนี้

$$\% Failure = P_{maintenance} \times P_{operation} \times 100$$



รูปที่ 4.12 ตัวอย่างการแจ้งเตือนค่าความผิดปกติมากกว่า 20%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.12 จะพบว่าการใช้งานความน่าจะเป็นของความผิดพลาดร่วมกัน ถ้าค่าความผิดพลาดมีค่ามากกว่า 20% จะทำการบันทึกและแจ้งเตือนเพื่อใช้ในการวางแผนดำเนินงานต่อไป

4.4 การอภิปรายผล

ในงานวิจัยนี้รูปแบบการแจกแจงความผิดพลาดจากข้อมูลการบำรุงรักษา พบว่าวิธีการแจกแจงผสมพหุนาม ให้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยกว่าวิธีการแจกแจงผสมเชิงเส้นและวิธีแจกแจงเชิงประกอบ อาจเนื่องมาจากรูปแบบความสัมพันธ์ระหว่างความน่าจะเป็นสะสมของความผิดพลาดแต่ละรูปแบบไม่เป็นเชิงเส้นกับความน่าจะเป็นสะสมของความผิดพลาดทั้งหมด นอกจากนี้วิธีแจกแจงผสมเชิงเส้นให้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำกว่าวิธีการแจกแจงเชิงประกอบ ซึ่งให้ผลตรงกันข้ามกับการศึกษาของ Roy et al. (2021) ซึ่งพบว่าอาจจะขึ้นอยู่กับข้อมูลที่พิจารณาหรือยังขึ้นอยู่กับปัจจัยอื่น เช่น การแบ่งกลุ่มของข้อมูลในแต่ละรูปแบบความผิดพลาด เป็นต้น

ส่วนความน่าจะเป็นของความผิดพลาดจากข้อมูลการใช้งานเครื่องจักรที่ไม่สัมพันธ์กับความเร็วรอบของเครื่องจักรวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวมีค่าคะแนน F1 และค่าเรียกคืนสูงกว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวไม่โดดเด่น ซึ่งไม่สอดคล้องกับการศึกษาของ Cetin and Tasgin (2021) ที่พบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวให้ผลดีกว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร เนื่องจากจากรูปแบบข้อมูลที่ผิดพลาดสามารถแยกออกได้อย่างชัดเจน จึงไม่จำเป็นต้องใช้วิธีที่มีความซับซ้อน

นอกจากนี้ความน่าจะเป็นของความผิดพลาดจากข้อมูลการใช้งานเครื่องจักรที่สัมพันธ์กับความเร็รรอบของเครื่องจักรวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานมีค่าคะแนน F1 ค่าความเที่ยง และค่าเรียกคืนสูงสุด และวิธีแผนภูมิควบคุมการถดถอยค่าคะแนน F1 ค่าความเที่ยง และค่าเรียกคืนสูงกว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร ซึ่งไม่สอดคล้องกับการศึกษาของ Cetin and Tasgin (2021) ที่พบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปรให้ผลดีกว่าวิธีส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร อาจเนื่องมาจากวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปรจะพิจารณาความผิดพลาดจากระยะห่างของข้อมูล ทำให้รูปแบบคะแนนความผิดพลาดเป็นลักษณะวงกลมซึ่งมีค่าคะแนนความผิดพลาดต่ำสุดบริเวณจุดศูนย์กลางวงกลมและลดลงตามระยะห่างจากจุดศูนย์กลาง จึงไม่เหมาะต่อการตรวจจับความผิดพลาดที่เป็นรูปแบบเชิงเส้น

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าความผิดพลาดของเครื่องจักรจากข้อมูลการบำรุงรักษาและข้อมูลการใช้งานเครื่องจักร ในข้อมูลการบำรุงรักษาเครื่องจักรจะใช้การแจกแจงไวบูลจำนวน 4 วิธี คือ วิธีไวบูล วิธีไวบูลเชิงประกอบ วิธีไวบูลผสมเชิงเส้น และวิธีไวบูลผสมพหุนาม โดยพิจารณาค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบ จากการศึกษาพบว่าวิธีไวบูลผสมพหุนามเป็นวิธีที่ดีที่สุดสำหรับข้อมูลที่กำหนด รองลงมาเป็นวิธีไวบูลเชิงประกอบ

ส่วนข้อมูลการใช้งานเครื่องจักรจะใช้การประมาณค่าความผิดพลาดเป็น 2 รูปแบบคือ ความผิดพลาดที่ไม่สัมพันธ์กันความเร็วของเครื่องจักร และความผิดพลาดที่สัมพันธ์กับความเร็วของเครื่องจักร ความผิดพลาดที่ไม่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดพลาดจาก 2 วิธี คือ วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวและวิธีป่าไม้โตดเดี่ยว จากการศึกษาพบว่าวิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานตัวแปรเดียวเป็นวิธีที่ดีที่สุด ในความผิดพลาดที่สัมพันธ์กับความเร็วของเครื่องจักรจะประมาณค่าความผิดพลาดจาก 3 วิธี คือ วิธีแผนภูมิถดถอย วิธี k เท่าของส่วนเบี่ยงเบนมาตรฐานหลายตัวแปร และวิธี k เท่าของส่วนเหลือมาตรฐาน พบว่า วิธี k เท่าของส่วนเหลือมาตรฐานเป็นวิธีที่ดีที่สุด โดยที่การประมาณค่าความผิดพลาดทั้ง 2 รูปแบบจะถูกทดสอบด้วยค่าความเที่ยง ค่าเรียกคืน และ คะแนน F1

5.2 ข้อเสนอแนะ

5.2.1 ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. ผู้สนใจทางด้านความเชื่อถือได้ของเครื่องจักรและอุปกรณ์สามารถนำผลการวิจัยนี้ไปใช้ประโยชน์ในการวางแผนงานวิจัยได้
2. ผู้วิจัยมีความตั้งใจให้แนวทางการวิจัยนี้ได้ถูกนำไปใช้งานจริงในภาคอุตสาหกรรม เพื่อนำผลการวิจัยนี้ไปประยุกต์ใช้กับเครื่องจักรและอุปกรณ์ได้จริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.2 ข้อเสนอแนะเพื่อเป็นแนวทางให้ผู้สนใจศึกษาเพิ่มเติม

1. ข้อมูลการบำรุงรักษาในงานวิจัยนี้เป็นข้อมูลย้อนหลังตั้งแต่วันที่ 1 พ.ค. 2563 ถึง 20 เม.ย. 2565 พบว่าเมื่อแบ่งตามรูปแบบความผิดปกติแล้วจำนวนข้อมูลในแต่ละรูปแบบมีจำนวนน้อย ควรกำหนดจำนวนข้อมูลตัวอย่างแบ่งตามรูปแบบความผิดปกติเพิ่มเติมเพื่อให้การคำนวณการแจกแจงปกติมีความแม่นยำขึ้น

2. ข้อมูลการใช้งานเครื่องจักรในงานวิจัยนี้มีเพียงข้อมูลการใช้งานเครื่องจักรในสภาวะปกติเท่านั้น ซึ่งในการทดสอบตัวแบบจำเป็นต้องสังเคราะห์ข้อมูลการใช้งานเครื่องจักรในสภาวะผิดปกติเพิ่มเติม ควรกำหนดจำนวนข้อมูลตัวอย่างทั้งสภาวะปกติและสภาวะผิดปกติเพิ่มเติมเพื่อให้การทดสอบมีประสิทธิภาพดีขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- Jain, R.C. and Anil, K. 1948. **Algorithm for Clustering Data**. New Jersey : Prentice Hall.
- Liu, D. H. F. and Tatera, J. F. 2006. **Advanced Topics in Statistical Process Control**. United States of America : CRC Press
- Fei, T. Ting, K.M. and Zhou, Z. "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*. 17 : 413-422.
<https://doi.org/10.1109/ICDM.2008.17>
- Pingwu, K. and Dewang, S. 2009. "Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space." *Pattern Recognition*. 42(5) : 710-717.
<https://doi.org/10.1016/j.patcog.2008.08.030>
- นิเวศ จิระวิฑิตชัย. 2010. **การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ**. กรุงเทพฯ : มหาวิทยาลัยราชภัฏสวนสุนันทา
<http://ssruir.ssru.ac.th/handle/ssruir/377>
- Kurz, Rainer Lubomirsky, Matt Brun and Klaus 2012. "Gas Compressor Station Economic Optimization." *International Journal of Rotating Machinery*. 715017 : 1-9. <https://doi.org/10.1155/2012/715017>
- Douglas, C.M. Elizabeth, A.P. and Geoffrey, G.V. 2013. **Introduction to Linear Regression Analysis**. 5rd ed. New Jersey : John Wiley & Sons.
- Douglas, C.M. and George, C.R. 2016. **Engineering Statistics**. 3rd ed. Bangkok : Top Publishing.
- Zhang, Y. Xiong, R. He, H. and Pecht, M.G. 2019. "Lithium-Ion battery remaining useful life prediction with Box-Cox transformation and Monte Carlo simulation" *IEEE Transactions on industrial electronic*. 66(2) : 1585-1597.
<https://doi.org/10.1109/TIE.2018.2808918>
- Condance, F. and Ronald L.J. 2019. "Principle assumptions of regression analysis: testing, techniques, and statistical reporting of imperfect data sets" *Advances in developing human resources*. 21(4) : 484-502.
<http://dx.doi.org/10.1177/1523422319869915>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

- Leukel, J. González, J. and Riekert, M. 2021. “Adoption of Machine Learning Technology for Failure Prediction in Industrial Maintenance: A Systematic Review.” *Journal of Manufacturing Systems*. 61 : 87-96.
<https://doi.org/10.1016/j.jmsy.2021.08.012>
- Zonta, T. Da Costa, C.A. Da Rosa Righi, R. De Lima, MJ. and Da Trindade, ES. 2020. “Predictive Maintenance in the Industry 4.0: A Systematic Literature Review.” *Computers and Industrial Engineering*. 150.
<https://doi.org/10.1016/j.cie.2020.106889>
- Montero Jimenez, J.J. Schwartz, S. Vingerhoeds, R. Grabot, B. and Salaün, M. 2020. “Towards Multi-Model Approaches to Predictive Maintenance: A Systematic Literature Survey on Diagnostics and Prognostics.” *Journal of Manufacturing Systems*. 56 : 539-557. <https://doi.org/10.1016/j.jmsy.2020.07.008>
- Lei, Y. Yang, B. Jiang, X. Jia, F. Li, N. and Nandi, A.K. 2020. “Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap.” *Journal of Manufacturing Systems*. 138. <https://doi.org/10.1016/j.ymsp.2019.106587>
- Alfeo, A.L. Cimino, M.G. and Vaglini, G. 2022. “Degradation Stage Classification via Interpretable Feature Learning.” *Journal of Manufacturing Systems*. 62 : 972–983. <https://doi.org/10.1016/j.jmsy.2021.05.003>
- De Jonge, B. Teunter, R. and Tinga, T. 2017. “The Influence of Practical Factors on the Benefits of Condition-Based Maintenance over Time-Based Maintenance.” *Reliability Engineering System Safety*. 158 : 21-30.
<https://doi.org/10.1016/j.res.2016.10.002>
- Kusiak, A. Verma, A. 2012. “A Data-Mining Approach to Monitoring Wind Turbines.” *IEEE Transactions on Sustainable Energy*. 3(1) : 150-157.
<https://doi.org/10.1109/TSTE.2011.2163177>
- Fernández, A. Vázquez, M. 2012 “Improved Estimation of Weibull Parameters Considering Unreliability Uncertainties.” *IEEE Transactions on Sustainable Energy*. 61(1) : 32-40. <https://doi.org/10.1109/TR.2011.2168652>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

- Jiang, X. 2014. "Multivariate Probabilistic Modelling for Seepage Risk Assessment in Tunnel Segments." *International Journal of Reliability and Safety*. 8(2) : 228-249. <http://dx.doi.org/10.1504/IJRS.2014.069507>
- Hayati, F. 2017. "Regression Control Chart for Two Related Variables: a Forgotten Lesson." *International Journal of Modelling in Operations Management*. 6(4) : 262-279. <http://dx.doi.org/10.1504/IJMOM.2017.10011283>
- Wang, J. and Yin, H. 2019. "Failure Rate Prediction Model of Substation Equipment Based on Weibull Distribution and Time Series Analysis." *IEEE Access*. 7 : 85298-85309. <https://doi.org/10.1109/ACCESS.2019.2926159>
- Yenal, U. and Jimenez, PD.. 2020. "Life data analysis with a joint probability density function." *Annual Reliability and Maintainability Symposium (RAMS)*. 1 : 1-6 <https://doi.org/10.1109/RAMS48030.2020.9153681>
- Roy, B. Bera D. Nigam, S. and Upadhyay, SK. 2021. "A Study of Turbine Failure Pattern: A Model Optimization Using Machine Learning." *International Journal of System Assurance Engineering and Management*. 13(4) : 1761-1770. <https://doi.org/10.1007/s13198-021-01542-9>
- Cetin, U. and Tasgin, M. 2021. "Anomaly Detection with Multivariate K-sigma Score Using Monte Carlo Sampling." *International Conference on Computer Science and Engineering (UBMK) 5th*. 1 : 94-98. <https://doi.org/10.1109/UBMK50275.2020.9219482>
- Das, K. 2016. "A brief review of tests for normality." *American journal of theoretical and applied statistics*. 5(1) : 5-12. <http://dx.doi.org/10.11648/j.ajtas.20160501.12>
- Guyon, I. and Elisseeff, A. 2003. "An introduction to variable and feature selection." *Journal of Machine Learning Research*. 3 : 1157-1182. <https://dl.acm.org/doi/10.5555/944919.944968>
- Rainer, K. Matt, L. and Klaus, B. 2021. "Gas Compressor Station Economic Optimization." *International Journal of Rotating Machinery*. 715017 : 1-9. <https://doi.org/10.1155/2012/715017>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Małgorzata, J. Kaczmarek. and A, Gola. 2019. “Maintenance 4.0 Technologies for Sustainable Manufacturing – an Overview.” *International Federation of Automatic Control*. 52(10): 91–96.

<https://doi.org/10.1016/j.ifacol.2019.10.005>.

Davide, C. Matthijs, J.W. and Giuseppe J. 2021. “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.” *PeerJ Computer Science*. 7(623)

<https://doi.org/10.7717/peerj-cs.623>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

โปรแกรมภาษาไพธอนแสดงขั้นตอนวิเคราะห์ข้อมูลการบำรุงรักษาเครื่องจักร

```
In [1]: # Data management
import pandas as pd
import numpy as np

# Visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Statistic and machine learning
import math
import statsmodels.api as sm
import scipy.stats as stats
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.ensemble import RandomForestRegressor
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
```

Maintenace Data Preparation

```
In [2]: # get the data and change data type
PM_CM_Data = pd.read_excel("../dataset/PM_CM_Data.xlsx")
PM_CM_Data['Basic_StartDate'] = pd.to_datetime(PM_CM_Data['Basic_StartDate'])
PM_CM_Data
```

```
Out[2]:
```

	EquipmentID	TagName	Main FL	EquipmentName	Model	SerialNo	Manufacturer	Equipment Type Code	Equipment Type Name	INGRP	...	DueDate	Act
0	2947	TSO-AFTERC	P-R05-RCS	After Cooler Skid	NaN	NaN	NaN	-	-	05C	...	NaT	
1	2947	TSO-AFTERC	P-R05-RCS	After Cooler Skid	NaN	NaN	NaN	-	-	05C	...	NaT	
2	2947	TSO-AFTERC	P-R05-RCS	After Cooler Skid	NaN	NaN	NaN	-	-	05C	...	NaT	
3	2947	TSO-AFTERC	P-R05-RCS	After Cooler Skid	NaN	NaN	NaN	-	-	05C	...	NaT	
4	2947	TSO-AFTERC	P-R05-RCS	After Cooler Skid	NaN	NaN	NaN	-	-	05C	...	NaT	

```
In [3]: # cleansing data and selection
mtdf = PM_CM_Data[~PM_CM_Data.Basic_StartDate.isna() & ~PM_CM_Data.OrderNo.isna()]
mtdf = mtdf.sort_values(['TagName', 'Basic_StartDate'], ascending=False)[['TagName', 'Main FL', 'Operation Short te
mtdf.columns = ['TagName', 'MainFL', 'OperationShortText', 'OrderNo', 'BasicStartDate', 'WorkType', 'FailureMode
mtdf['FailureMode'] = np.where(~mtdf['FailureMode'].isna(), mtdf['FailureMode'], mtdf['OperationShortText'])
mtdf['FL_FM'] = mtdf.MainFL.map(str) + "_" + mtdf.TagName + "_" + mtdf['FailureMode']
mtdf = mtdf.drop(['OperationShortText', 'TagName'], axis=1)
mtdf = mtdf.drop_duplicates(subset='OrderNo')
mtdf
```

```
Out[3]:
```

	MainFL	OrderNo	BasicStartDate	WorkType	FailureMode	FL_FM
2862	TSO-YARD_HD	120764202.0	2022-04-01	PM	BDVML3:5Y	TSO-YARD_HD_TSO-YARD_HD-4200-BDV-417_BDVML3:5Y
2861	TSO-YARD_HD	120718056.0	2021-09-01	PM	BDVML2:1Y	TSO-YARD_HD_TSO-YARD_HD-4200-BDV-417_BDVML2:1Y
2860	TSO-YARD_HD	120656111.0	2020-12-01	PM	BDVML2:1Y	TSO-YARD_HD_TSO-YARD_HD-4200-BDV-417_BDVML2:1Y
2859	TSO-YARD_HD	120635198.0	2020-09-01	PM	Blow Down Valve	TSO-YARD_HD_TSO-YARD_HD-4200-BDV-417_Blow Down...
2858	TSO-YARD_HD	120613970.0	2020-06-01	PM	Blow Down Valve	TSO-YARD_HD_TSO-YARD_HD-4200-BDV-417_Blow Down...

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Gas Turbine

```
In [4]: # focus on gas turbine data
gtdf = mtdf[mtdf.MainFL.str.contains('ENGIN')].sort_values(['FL_FM', 'BasicStartDate']).reset_index(drop=True)
gtcdf
```

```
Out[4]:
```

	MainFL	OrderNo	BasicStartDate	WorkType	FailureMode	FL_FM
0	TSO-ENGIN_A	120680021.0	2021-01-01	PM	IGVML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-EGV-943A_IGVML2:1Y
1	TSO-ENGIN_A	120743744.0	2022-01-01	PM	IGVML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-EGV-943A_IGVML2:1Y
2	TSO-ENGIN_A	120652663.0	2020-11-01	PM	GTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y
3	TSO-ENGIN_A	120720424.0	2021-10-01	PM	GTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y
4	TSO-ENGIN_A	120774781.0	2022-09-01	PM	GTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y
5	TSO-ENGIN_A	110249360.0	2022-10-28	CM	Other	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_Other
6	TSO-ENGIN_A	120680050.0	2021-01-01	PM	TTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-TE-867A_TTML2:1Y
7	TSO-ENGIN_A	120743138.0	2022-01-01	PM	TTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-TE-867A_TTML2:1Y
8	TSO-ENGIN_A	120774750.0	2022-04-20	PM	VTPML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-YT-859A_VTPML2:1Y
9	TSO-ENGIN_B	120607016.0	2020-05-01	PM	IGVML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y
10	TSO-ENGIN_B	120683910.0	2021-02-01	PM	IGVML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y
11	TSO-ENGIN_B	120748065.0	2022-02-01	PM	IGVML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y

```
In [5]: # calculate time to fail
gtcdf['TTF'] = (gtcdf.BasicStartDate-gtcdf.shift(1).BasicStartDate).dt.days
gtcdf['TTF'] = np.where(gtcdf.FL_FM == gtcdf.shift(1).FL_FM,gtcdf['TTF'],0)
gtcdf_clean = gtcdf[gtcdf.TTF>0].sort_values(['FailureMode', 'TTF']).reset_index(drop=True)
gtcdf_clean
```

```
Out[5]:
```

	MainFL	OrderNo	BasicStartDate	WorkType	FailureMode	FL_FM	TTF
0	TSO-ENGIN_C	120720425.0	2021-11-01	PM	GTML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y	61.0
1	TSO-ENGIN_C	120720423.0	2021-09-01	PM	GTML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y	304.0
2	TSO-ENGIN_A	120720424.0	2021-10-01	PM	GTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y	334.0
3	TSO-ENGIN_C	120774783.0	2022-10-01	PM	GTML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y	334.0
4	TSO-ENGIN_A	120774781.0	2022-09-01	PM	GTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y	335.0
5	TSO-ENGIN_B	120774782.0	2022-10-01	PM	GTML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-GT-001B_GTML2:1Y	699.0
6	TSO-ENGIN_C	120690471.0	2021-03-01	PM	IGVML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943C_IGVML2:1Y	273.0
7	TSO-ENGIN_B	120683910.0	2021-02-01	PM	IGVML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y	276.0
8	TSO-ENGIN_A	120743744.0	2022-01-01	PM	IGVML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-EGV-943A_IGVML2:1Y	365.0
9	TSO-ENGIN_B	120748065.0	2022-02-01	PM	IGVML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y	365.0
10	TSO-ENGIN_C	120757680.0	2022-03-01	PM	IGVML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943C_IGVML2:1Y	365.0
11	TSO-ENGIN_C	120690478.0	2021-03-01	PM	TTML2:1Y	TSO-ENGIN_C_TSO-ENGIN_C-4200-TE-867C_TTML2:1Y	273.0
12	TSO-ENGIN_B	120683911.0	2021-02-01	PM	TTML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-TE-867B_TTML2:1Y	276.0
13	TSO-ENGIN_A	120743138.0	2022-01-01	PM	TTML2:1Y	TSO-ENGIN_A_TSO-ENGIN_A-4200-TE-867A_TTML2:1Y	365.0
14	TSO-ENGIN_B	120749175.0	2022-02-01	PM	TTML2:1Y	TSO-ENGIN_B_TSO-ENGIN_B-4200-TE-867B_TTML2:1Y	365.0

```
In [6]: # define function calculate CDF value calculation form benard equation
def getFt (n):
df = pd.DataFrame(range(1,n+1),columns=['n'])
df['Ft'] = (df.n-0.3)/(n+0.4)
return df['Ft']
```

```
In [7]: # define function finding the weibull distribution parameter
def getweibparam (df):
df = df.sort_values('TTF')
x = np.log(df.TTF)
y = np.log(np.log(1/(1-df.Ft)))
param = stats.linregress(x,y)
shape = param.slope
scale = np.exp(param.intercept/-param.slope)
return [shape,scale]
```

เอกสารนี้เป็นเอกสารที่สวชนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [8]: # sort time to fail value and calculate the CDF value
gtdf_clean_all = gtdf_clean.copy()
gtdf_clean_all = gtdf_clean_all.sort_values('TTF').reset_index(drop=True)
gtdf_clean_all['Ft'] = getFt(len(gtdf_clean_all))
gtdf_clean_all
```

```
Out[8]:
```

	MainFL	OrderNo	BasicStartDate	WorkType	FailureMode	FL_FM	TTF	Ft
0	TSO-ENGIN_C	120720425.0	2021-11-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		61.0	0.045455
1	TSO-ENGIN_C	120690471.0	2021-03-01	PM	IGVML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943B_IGVML2:1Y		273.0	0.110390
2	TSO-ENGIN_C	120690478.0	2021-03-01	PM	TTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-TE-867C_TTML2:1Y		273.0	0.175325
3	TSO-ENGIN_B	120683910.0	2021-02-01	PM	IGVML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y		276.0	0.240260
4	TSO-ENGIN_B	120683911.0	2021-02-01	PM	TTML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-TE-867B_TTML2:1Y		276.0	0.305195
5	TSO-ENGIN_C	120720423.0	2021-09-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		304.0	0.370130
6	TSO-ENGIN_A	120720424.0	2021-10-01	PM	GTML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y		334.0	0.435065
7	TSO-ENGIN_C	120774783.0	2022-10-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		334.0	0.500000
8	TSO-ENGIN_A	120774781.0	2022-09-01	PM	GTML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y		335.0	0.564935
9	TSO-ENGIN_A	120743744.0	2022-01-01	PM	IGVML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-EGV-943A_IGVML2:1Y		365.0	0.629870
10	TSO-ENGIN_B	120748065.0	2022-02-01	PM	IGVML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y		365.0	0.694805
11	TSO-ENGIN_C	120757680.0	2022-03-01	PM	IGVML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943C_IGVML2:1Y		365.0	0.759740

Traditional weibull model

```
In [9]: # data selection
gtdf_clean_all[['MainFL', 'BasicStartDate', 'TTF', 'Ft']]
```

```
Out[9]:
```

	MainFL	BasicStartDate	TTF	Ft
0	TSO-ENGIN_C	2021-11-01	61.0	0.045455
1	TSO-ENGIN_C	2021-03-01	273.0	0.110390
2	TSO-ENGIN_C	2021-03-01	273.0	0.175325
3	TSO-ENGIN_B	2021-02-01	276.0	0.240260
4	TSO-ENGIN_B	2021-02-01	276.0	0.305195
5	TSO-ENGIN_C	2021-09-01	304.0	0.370130
6	TSO-ENGIN_A	2021-10-01	334.0	0.435065
7	TSO-ENGIN_C	2022-10-01	334.0	0.500000
8	TSO-ENGIN_A	2022-09-01	335.0	0.564935
9	TSO-ENGIN_A	2022-01-01	365.0	0.629870
10	TSO-ENGIN_B	2022-02-01	365.0	0.694805
11	TSO-ENGIN_C	2022-03-01	365.0	0.759740
12	TSO-ENGIN_A	2022-01-01	365.0	0.824675
13	TSO-ENGIN_B	2022-02-01	365.0	0.889610
14	TSO-ENGIN_B	2022-10-01	699.0	0.954545

```
In [10]: # calculate weibull parameter [shape, scale]
gtdf_clean_all_weibparam = getweibparam(gtdf_clean_all)
gtdf_clean_all_weibparam
```

```
Out[10]: [1.9215281641517281, 402.4355487046203]
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Weibull distribution in each failure mode

```
In [11]: gtdf_clean_fm = gtdf_clean.copy()
```

```
In [12]: # Calculate CDF value in each failure mode
tempdf_list = list()
for i in gtdf_clean_fm.FailureMode.unique():
    tempdf = gtdf_clean_fm[gtdf_clean_fm.FailureMode==i].copy()
    tempdf['Ft'] = getFt(len(tempdf)).values
    tempdf_list.append(tempdf)
gtdf_clean_fm = pd.concat(tempdf_list)
gtdf_clean_fm
```

```
Out[12]:
```

	MainFL	OrderNo	BasicStartDate	WorkType	FailureMode	FL_FM	TTF	Ft
0	TSO-ENGIN_C	120720425.0	2021-11-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		61.0	0.109375
1	TSO-ENGIN_C	120720423.0	2021-09-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		304.0	0.265625
2	TSO-ENGIN_A	120720424.0	2021-10-01	PM	GTML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y		334.0	0.421875
3	TSO-ENGIN_C	120774783.0	2022-10-01	PM	GTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-GT-001C_GTML2:1Y		334.0	0.578125
4	TSO-ENGIN_A	120774781.0	2022-09-01	PM	GTML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-GT-001A_GTML2:1Y		335.0	0.734375
5	TSO-ENGIN_B	120774782.0	2022-10-01	PM	GTML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-GT-001B_GTML2:1Y		699.0	0.890625
6	TSO-ENGIN_C	120690471.0	2021-03-01	PM	IGVML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943C_IGVML2:1Y		273.0	0.129630
7	TSO-ENGIN_B	120683910.0	2021-02-01	PM	IGVML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y		276.0	0.314815
8	TSO-ENGIN_A	120743744.0	2022-01-01	PM	IGVML2:1Y TSO-ENGIN_A_TSO-ENGIN_A-4200-EGV-943A_IGVML2:1Y		365.0	0.500000
9	TSO-ENGIN_B	120748065.0	2022-02-01	PM	IGVML2:1Y TSO-ENGIN_B_TSO-ENGIN_B-4200-EGV-943B_IGVML2:1Y		365.0	0.685185
10	TSO-ENGIN_C	120757680.0	2022-03-01	PM	IGVML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-EGV-943C_IGVML2:1Y		365.0	0.870370
11	TSO-ENGIN_C	120690478.0	2021-03-01	PM	TTML2:1Y TSO-ENGIN_C_TSO-ENGIN_C-4200-TE-867C_TTML2:1Y		273.0	0.159091

```
In [13]: # Calculate Weibull parameter [shape, scale] in each failure mode
gtdf_clean_fm_weibparam = pd.DataFrame(columns=['FailureMode', 'weibparam'])
for i in gtdf_clean_fm.FailureMode.unique():
    tempdf = gtdf_clean_fm[gtdf_clean_fm.FailureMode==i]
    weibparam = getweibparam(tempdf)
    gtdf_clean_fm_weibparam = gtdf_clean_fm_weibparam.append({'FailureMode': i, 'weibparam': weibparam}, ignore_index=True)
gtdf_clean_fm_weibparam
```

```
Out[13]:
```

	FailureMode	weibparam
0	GTML2:1Y	[1.1765851769210385, 428.8511415825653]
1	IGVML2:1Y	[5.80010819318955, 354.4428254941219]
2	TTML2:1Y	[5.379088446447412, 345.9295764141948]

```
In [14]: # define weibull CDF and PDF function
weibull_cdf = lambda x, shape, scale : 1-np.exp(-(x/scale)**shape)
weibull_pdf = lambda x, shape, scale : ((shape*x**(shape-1))*np.exp(-(x/scale)**shape))/scale**shape
```

Analysis base table

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

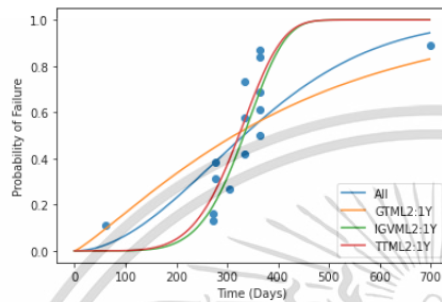
```
In [15]: # plot weibull CDF comparing between traditional (all failure mode) and each failure distribution
plt.scatter(gtdf_clean_fm.TTF,gtdf_clean_fm.Ft)

x = [i for i in range(int(gtdf_clean_fm.TTF.max()))]

cdf = weibull_cdf(x,gtdf_clean_all_weibparam[0],gtdf_clean_all_weibparam[1])
plt.plot(x,cdf,label='All')

for i in gtdf_clean_fm_weibparam.FailureMode:
    shape = gtdf_clean_fm_weibparam[gtdf_clean_fm_weibparam.FailureMode==i].weibparam.values[0][0]
    scale = gtdf_clean_fm_weibparam[gtdf_clean_fm_weibparam.FailureMode==i].weibparam.values[0][1]
    cdf = weibull_cdf(x,shape,scale)
    plt.plot(x,cdf,label=i)

plt.legend(loc='lower right')
plt.xlabel('Time (Days)')
plt.ylabel('Probability of Failure')
plt.show()
```



```
In [16]: # Data selection for build model
gtdf_clean_abt = pd.DataFrame()
gtdf_clean_abt['TTF'] = gtdf_clean_all.TTF
gtdf_clean_abt['Ft'] = gtdf_clean_all.Ft
gtdf_clean_abt['All'] = weibull_cdf(gtdf_clean_all.TTF,gtdf_clean_all_weibparam[0],gtdf_clean_all_weibparam[1])
for i in gtdf_clean_fm_weibparam.FailureMode:
    shape = gtdf_clean_fm_weibparam[gtdf_clean_fm_weibparam.FailureMode==i].weibparam.values[0][0]
    scale = gtdf_clean_fm_weibparam[gtdf_clean_fm_weibparam.FailureMode==i].weibparam.values[0][1]
    gtdf_clean_abt[i] = weibull_cdf(gtdf_clean_all.TTF,shape,scale)
gtdf_clean_abt
```

```
Out[16]:
```

	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

composite distribution

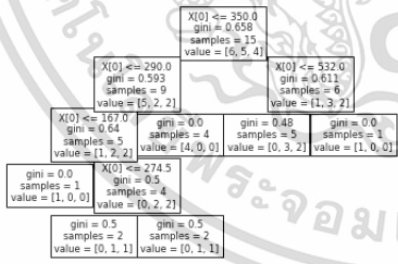
```
In [17]: # calculate absolute error in each CDF failure mode compare with benard
gtdf_clean_abt['GTML2_error'] = (gtdf_clean_abt['GTML2:1Y']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt['IGVML2_error'] = (gtdf_clean_abt['IGVML2:1Y']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt['TTML2_error'] = (gtdf_clean_abt['TTML2:1Y']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt['All_error'] = (gtdf_clean_abt['All']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt
```

Out[17]:	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y	GTML2_error	IGVML2_error	TTML2_error	All_error
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088	0.050431	0.045418	0.045366	0.019164
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088	0.334055	0.087067	0.133698	0.267361
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088	0.269120	0.022132	0.068763	0.202426
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788	0.208395	0.031333	0.016529	0.143724
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788	0.143460	0.096268	0.048406	0.078789
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902	0.116666	0.033462	0.022772	0.071833
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066	0.090293	0.072568	0.128001	0.067833
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066	0.025358	0.007633	0.063066	0.002898
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891	0.038332	0.051238	0.003956	0.060039
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744	0.067132	0.064562	0.106874	0.066384
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744	0.132067	0.000373	0.041939	0.131319
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744	0.197002	0.065308	0.022996	0.196254
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744	0.261937	0.130243	0.087931	0.261189
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744	0.326872	0.195178	0.152866	0.326124
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000	0.123725	0.045455	0.045455	0.010178

```
In [18]: # Prepare data to find a separation point of composite distribution
X = np.array(gtfd_clean_fm.TTF).reshape(-1, 1)
y = gtfd_clean_fm.FailureMode
```

```
In [19]: # Fitting Tree
dtree = DecisionTreeClassifier()
dtree = dtree.fit(X, y)
tree.plot_tree(dtree)
```

```
Out[19]: [Text(0.5555555555555556, 0.9, 'X[0] <= 350.0\nngini = 0.658\nnsamples = 15\nnvalue = [6, 5, 4]'),
Text(0.3333333333333333, 0.7, 'X[0] <= 290.0\nngini = 0.593\nnsamples = 9\nnvalue = [5, 2, 2]'),
Text(0.2222222222222222, 0.5, 'X[0] <= 167.0\nngini = 0.64\nnsamples = 5\nnvalue = [1, 2, 2]'),
Text(0.1111111111111111, 0.3, 'gini = 0.0\nnsamples = 1\nnvalue = [1, 0, 0]'),
Text(0.3333333333333333, 0.3, 'X[0] <= 274.5\nngini = 0.5\nnsamples = 4\nnvalue = [0, 2, 2]'),
Text(0.2222222222222222, 0.1, 'gini = 0.5\nnsamples = 2\nnvalue = [0, 1, 1]'),
Text(0.4444444444444444, 0.1, 'gini = 0.5\nnsamples = 2\nnvalue = [0, 1, 1]'),
Text(0.4444444444444444, 0.5, 'gini = 0.0\nnsamples = 4\nnvalue = [4, 0, 0]'),
Text(0.7777777777777778, 0.7, 'X[0] <= 532.0\nngini = 0.611\nnsamples = 6\nnvalue = [1, 3, 2]'),
Text(0.6666666666666666, 0.5, 'gini = 0.48\nnsamples = 5\nnvalue = [0, 3, 2]'),
Text(0.8888888888888888, 0.5, 'gini = 0.0\nnsamples = 1\nnvalue = [1, 0, 0]')]
```



เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [20]: # select nearest CDF value in each failure mode compare with benard
gtddf_clean_abt['composite'] = np.where((gtddf_clean_abt['TTF']>350),gtddf_clean_abt['TTML2:1Y'],gtddf_clean_abt['IGV
gtddf_clean_abt
```

```
Out[20]:
```

	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y	GTML2_error	IGVML2_error	TTML2_error	All_error	composite
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088	0.050431	0.045418	0.045366	0.019164	0.000037
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088	0.334055	0.087067	0.133698	0.267361	0.197457
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088	0.269120	0.022132	0.068763	0.202426	0.197457
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788	0.208395	0.031333	0.016529	0.143724	0.208927
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788	0.143460	0.096268	0.048406	0.078789	0.208927
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902	0.116666	0.033462	0.022772	0.071833	0.336668
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066	0.090293	0.072568	0.128001	0.067833	0.507633
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066	0.025358	0.007633	0.063066	0.002898	0.507633
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891	0.038332	0.051238	0.003956	0.060039	0.513697
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744	0.067132	0.064562	0.106874	0.066384	0.736744
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744	0.132067	0.000373	0.041939	0.131319	0.736744
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744	0.197002	0.065308	0.022996	0.196254	0.736744
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744	0.261937	0.130243	0.087931	0.261189	0.736744
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744	0.326872	0.195178	0.152866	0.326124	0.736744
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000	0.123725	0.045455	0.045455	0.010178	1.000000

```
In [21]: # calculate composite distribution absolute error
gtddf_clean_abt['composite_error'] = (gtddf_clean_abt['composite']-gtddf_clean_abt.Ft).abs()
gtddf_clean_abt
```

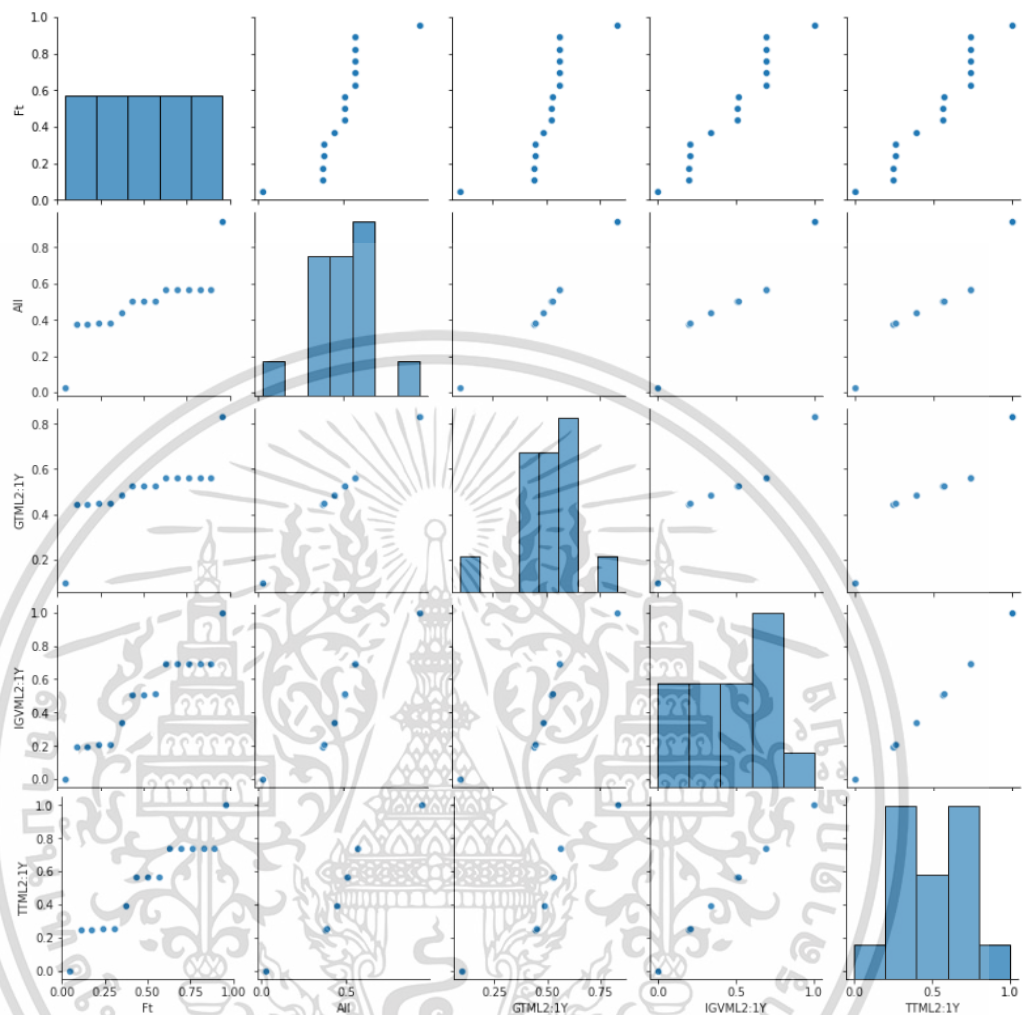
```
Out[21]:
```

	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y	GTML2_error	IGVML2_error	TTML2_error	All_error	composite	composite_er
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088	0.050431	0.045418	0.045366	0.019164	0.000037	0.045455
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088	0.334055	0.087067	0.133698	0.267361	0.197457	0.087067
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088	0.269120	0.022132	0.068763	0.202426	0.197457	0.022132
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788	0.208395	0.031333	0.016529	0.143724	0.208927	0.031333
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788	0.143460	0.096268	0.048406	0.078789	0.208927	0.096268
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902	0.116666	0.033462	0.022772	0.071833	0.336668	0.033462
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066	0.090293	0.072568	0.128001	0.067833	0.507633	0.072568
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066	0.025358	0.007633	0.063066	0.002898	0.507633	0.007633
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891	0.038332	0.051238	0.003956	0.060039	0.513697	0.051238
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744	0.067132	0.064562	0.106874	0.066384	0.736744	0.106874
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744	0.132067	0.000373	0.041939	0.131319	0.736744	0.041939
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744	0.197002	0.065308	0.022996	0.196254	0.736744	0.022996
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744	0.261937	0.130243	0.087931	0.261189	0.736744	0.087931
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744	0.326872	0.195178	0.152866	0.326124	0.736744	0.152866
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000	0.123725	0.045455	0.045455	0.010178	1.000000	0.045455

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mix distribution

```
In [22]: # visual relation between CDF value in each failure mode compare with benard
sns.pairplot(gtdf_clean_abt[['Ft', 'All', 'GTML2:1Y', 'IGVML2:1Y', 'TTML2:1Y']])
Out[22]: <seaborn.axisgrid.PairGrid at 0x2bela0580>
```



```
In [23]: # prepare data for mix distribution
X = gtdf_clean_abt[['GTML2:1Y', 'IGVML2:1Y', 'TTML2:1Y']]
Y = gtdf_clean_abt[['Ft']]
```

```
In [24]: # build mixed distribution model by OLS regression
X_linear = sm.add_constant(X)
result1 = sm.OLS(Y, X_linear).fit()
```

```
In [25]: # confirm residual distribution as normal with JB testing
stats.jarque_bera(result1.resid)
```

```
Out[25]: Jarque_beraResult(statistic=0.3853339475526721, pvalue=0.8247565939827013)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [26]: # result of mixed distribution
gtdf_clean_abt['mix'] = result1.predict(X_linear)
gtdf_clean_abt
```

```
Out[26]:
```

	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y	GTML2_error	IGVML2_error	TTML2_error	All_error	composite	composite_er
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088	0.050431	0.045418	0.045366	0.019164	0.000037	0.0454
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088	0.334055	0.087067	0.133698	0.267361	0.197457	0.0870
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088	0.269120	0.022132	0.068763	0.202426	0.197457	0.0221
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788	0.208395	0.031333	0.016529	0.143724	0.208927	0.0313
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788	0.143460	0.096268	0.048406	0.078789	0.208927	0.0962
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902	0.116666	0.033462	0.022772	0.071833	0.336668	0.0334
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066	0.090293	0.072568	0.128001	0.067833	0.507633	0.0725
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066	0.025358	0.007633	0.063066	0.002898	0.507633	0.0076
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891	0.038332	0.051238	0.003956	0.060039	0.513697	0.0512
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744	0.067132	0.064562	0.106874	0.066384	0.736744	0.1068
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744	0.132067	0.000373	0.041939	0.131319	0.736744	0.0419
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744	0.197002	0.065308	0.022996	0.196254	0.736744	0.0229
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744	0.261937	0.130243	0.087931	0.261189	0.736744	0.0879
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744	0.326872	0.195178	0.152866	0.326124	0.736744	0.1528
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000	0.123725	0.045455	0.045455	0.010178	1.000000	0.0454

polynomial mixed distribution

```
In [27]: # prepare polynomial data (degree 2) for build model
poly = PolynomialFeatures(2)
X_poly2 = poly.fit_transform(X)
```

```
In [28]: # build model polynomial distribution by ols regression
result2 = sm.OLS(Y, X_poly2).fit()
result2.rsquared
```

```
Out[28]: 0.9589285714285715
```

```
In [29]: # confirm residual distribution as normal with JB testing
stats.jarque_bera(result2.resid)
```

```
Out[29]: Jarque_beraResult(statistic=0.5049748115894388, pvalue=0.7768659967875418)
```

```
In [30]: # prepare polynomial data (degree 3) for build model
poly = PolynomialFeatures(3)
X_poly3 = poly.fit_transform(X)
```

```
In [31]: # build model polynomial distribution by ols regression
result3 = sm.OLS(Y, X_poly3).fit()
result3.rsquared
```

```
Out[31]: 0.9589285714283463
```

```
In [32]: # confirm residual distribution as normal with JB testing
stats.jarque_bera(result3.resid)
```

```
Out[32]: Jarque_beraResult(statistic=0.5049748115896134, pvalue=0.776865996787474)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [33]: # Selected lowest degree of polynomial for mixed polynomial distribution
gtdf_clean_abt['poly'] = result2.predict(X_poly2)
```

```
In [34]: # calculate absolute error
gtdf_clean_abt['mix_error'] = (gtdf_clean_abt['mix']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt['poly_error'] = (gtdf_clean_abt['poly']-gtdf_clean_abt.Ft).abs()
gtdf_clean_abt
```

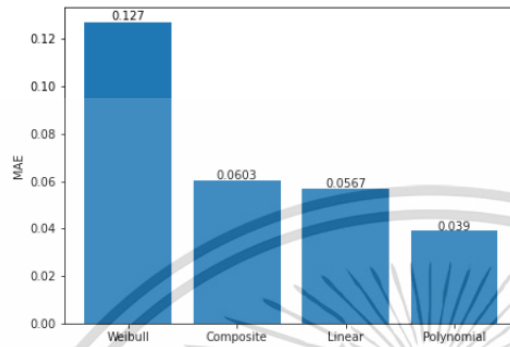
```
Out[34]:
```

	TTF	Ft	All	GTML2:1Y	IGVML2:1Y	TTML2:1Y	GTML2_error	IGVML2_error	TTML2_error	All_error	composite	composite_er
0	61.0	0.045455	0.026290	0.095886	0.000037	0.000088	0.050431	0.045418	0.045366	0.019164	0.000037	0.0454
1	273.0	0.110390	0.377751	0.444445	0.197457	0.244088	0.334055	0.087067	0.133698	0.267361	0.197457	0.0870
2	273.0	0.175325	0.377751	0.444445	0.197457	0.244088	0.269120	0.022132	0.068763	0.202426	0.197457	0.0221
3	276.0	0.240260	0.383984	0.448655	0.208927	0.256788	0.208395	0.031333	0.016529	0.143724	0.208927	0.0313
4	276.0	0.305195	0.383984	0.448655	0.208927	0.256788	0.143460	0.096268	0.048406	0.078789	0.208927	0.0962
5	304.0	0.370130	0.441963	0.486796	0.336668	0.392902	0.116666	0.033462	0.022772	0.071833	0.336668	0.0334
6	334.0	0.435065	0.502898	0.525358	0.507633	0.563066	0.090293	0.072568	0.128001	0.067833	0.507633	0.0725
7	334.0	0.500000	0.502898	0.525358	0.507633	0.563066	0.025358	0.007633	0.063066	0.002898	0.507633	0.0076
8	335.0	0.564935	0.504896	0.526603	0.513697	0.568891	0.038332	0.051238	0.003956	0.060039	0.513697	0.0512
9	365.0	0.629870	0.563486	0.562738	0.694432	0.736744	0.067132	0.064562	0.106874	0.066384	0.736744	0.1068
10	365.0	0.694805	0.563486	0.562738	0.694432	0.736744	0.132067	0.000373	0.041939	0.131319	0.736744	0.0419
11	365.0	0.759740	0.563486	0.562738	0.694432	0.736744	0.197002	0.065308	0.022996	0.196254	0.736744	0.0229
12	365.0	0.824675	0.563486	0.562738	0.694432	0.736744	0.261937	0.130243	0.087931	0.261189	0.736744	0.0879
13	365.0	0.889610	0.563486	0.562738	0.694432	0.736744	0.326872	0.195178	0.152866	0.326124	0.736744	0.1528
14	699.0	0.954545	0.944367	0.830820	1.000000	1.000000	0.123725	0.045455	0.045455	0.010178	1.000000	0.0454

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [35]: # calculate mean absolute error
def addlabels(x,y):
    for i in range(len(x)):
        plt.text(i, y[i]*1.01, round(y[i],4), ha = 'center')

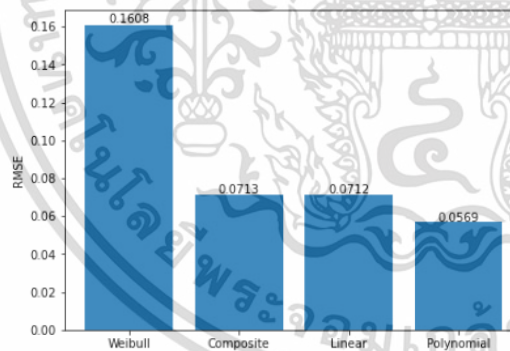
plt.figure(figsize=(7,5))
mae = gtdf_clean_abt[['All_error', 'composite_error','mix_error','poly_error']].mean(axis=0)
mae.index = ['Weibull', 'Composite', 'Linear', 'Polynomial']
plt.bar(mae.index,mae.values)
addlabels(mae.index,mae.values)
plt.ylabel('MAE')
plt.show()
```



```
In [36]: # calculate root mean square error
rmse = [np.sqrt(mean_squared_error(gtdf_clean_abt.All,gtdf_clean_abt.Ft)),
        np.sqrt(mean_squared_error(gtdf_clean_abt.composite,gtdf_clean_abt.Ft)),
        np.sqrt(mean_squared_error(gtdf_clean_abt.mix,gtdf_clean_abt.Ft)),
        np.sqrt(mean_squared_error(gtdf_clean_abt.poly,gtdf_clean_abt.Ft))]

def addlabels(x,y):
    for i in range(len(x)):
        plt.text(i, y[i]*1.01, round(y[i],4), ha = 'center')

plt.figure(figsize=(7,5))
mae.index = ['Weibull', 'Composite', 'Linear', 'Polynomial']
plt.bar(mae.index,rmse)
addlabels(mae.index,rmse)
plt.ylabel('RMSE')
plt.show()
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

โปรแกรมภาษาไพธอนแสดงขั้นตอนวิเคราะห์ข้อมูลการใช้งานเครื่องจักร

```
In [1]: # Data Mangement
import pandas as pd
import numpy as np

# Visualizaton
import matplotlib.pyplot as plt
import seaborn as sns

# Statistic and Machine Learning
import math
import statsmodels.api as sm
from scipy import stats
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import IsolationForest
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score, precision_score, recall_score
```

Classify relation and individual model

```
In [2]: # Get historian operation data
RCS_A = pd.read_csv('dataset/RCS_A_load.csv', index_col='PI Tag').I
RCS_A
```

```
Out[2]:
```

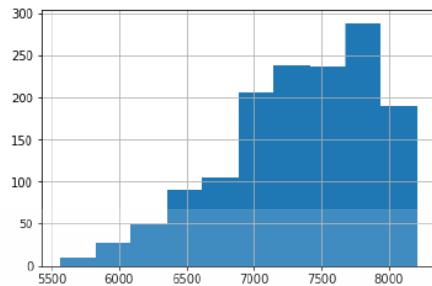
PI Tag	TSO-DGS_A-4200-PT-736A-PV	TSO-ENGIN_A-4200-ST-438-PV	TSO-ENGIN_A-4200-TE-851A-PV	TSO-ENGIN_A-4200-TE-852A-PV	TSO-ENGIN_A-4200-TE-853A-PV	TSO-ENGIN_A-4200-TE-854A-PV	TSO-ENGIN_A-4200-TE-855A-PV	TSO-ENGIN_A-4200-TE-856A-PV	TSO-ENGIN_A-4200-TE-857A-PV	TSO-ENGIN_A-4200-TE-858A-PV	TSO-ENGIN_A-4200-TE-942A-FB	TSO-ENGIN_J4200-EBV-942A-CM	
23/3/2022 0:00	0.50	92.47	1120.96	1156.05	1155.23	1132.86	1130.23	1144.80	1108.63	1134.21	...	0.03	0
23/3/2022 0:05	0.50	92.43	1117.23	1154.36	1151.83	1132.40	1129.86	1144.23	1105.98	1129.80	...	0.03	0
23/3/2022 0:10	0.50	92.39	1116.86	1153.15	1153.19	1131.61	1129.50	1145.70	1104.65	1128.86	...	0.03	0
23/3/2022 0:15	0.50	92.34	1115.00	1150.94	1152.08	1130.81	1128.93	1143.68	1103.32	1127.93	...	0.03	0
23/3/2022 0:20	0.50	92.30	1112.97	1150.72	1151.18	1130.01	1127.54	1140.25	1102.66	1129.24	...	0.03	0
...
27/3/2022 23:40	0.48	93.60	1144.20	1176.59	1183.49	1149.81	1147.82	1178.43	1136.04	1156.99	...	0.06	0
27/3/2022 23:45	0.48	93.32	1135.60	1166.10	1173.17	1140.22	1135.94	1167.33	1127.12	1149.82	...	0.05	0
27/3/2022 23:50	0.48	93.01	1126.16	1166.03	1171.87	1138.79	1133.30	1161.98	1124.18	1141.88	...	0.05	0
27/3/2022 23:55	0.48	92.84	1124.06	1166.56	1170.29	1136.70	1133.51	1159.63	1123.44	1144.16	...	0.05	0
28/3/2022 0:00	0.48	92.67	1127.74	1165.85	1171.51	1135.80	1133.73	1161.66	1122.46	1143.97	...	0.05	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [3]: # Define speed tag name for classify data that relate to speed
speedtag = 'AP_HPC_Spd_Act'
```

```
In [4]: # Check speed data distribution to filter only on-load speed period
RCS_A[speedtag].hist()
```

Out[4]: <AxesSubplot:>



```
In [5]: # cleansing data with on-load speed more than 7500 rpm and drop incorrect data
RCS_A = RCS_A[RCS_A[speedtag]>7500]
tag = pd.DataFrame(RCS_A.describe().loc['min'][:]>0)
RCS_A = RCS_A[tag[tag['min']==True].index]
RCS_A
```

Out[5]:

PI Tag	TSO-DGS_A-4200-PT-736A-PV	TSO-ENGIN_A-4200-ST-438-PV	TSO-ENGIN_A-4200-TE-851A-PV	TSO-ENGIN_A-4200-TE-852A-PV	TSO-ENGIN_A-4200-TE-853A-PV	TSO-ENGIN_A-4200-TE-854A-PV	TSO-ENGIN_A-4200-TE-855A-PV	TSO-ENGIN_A-4200-TE-856A-PV	TSO-ENGIN_A-4200-TE-857A-PV	TSO-ENGIN_A-4200-TE-858A-PV	TSO-ENCAN_A-4200-RT382-PV	TSO-ENCAN_A-4200-985A-PV	
23/3/2022 10:15	0.47	96.17	1176.35	1219.08	1223.05	1194.95	1193.72	1207.86	1173.05	1191.31	...	99.59	95
23/3/2022 10:20	0.47	96.25	1178.54	1222.55	1225.16	1197.35	1195.38	1209.87	1174.26	1192.65	...	99.49	95
23/3/2022 10:25	0.47	96.33	1178.27	1225.34	1225.60	1199.27	1197.09	1211.12	1173.18	1194.48	...	99.78	96
23/3/2022 10:30	0.47	96.41	1180.19	1228.05	1232.08	1201.00	1198.84	1215.07	1176.66	1193.98	...	100.05	96
23/3/2022 10:35	0.47	96.49	1181.30	1233.10	1232.13	1202.96	1201.54	1217.79	1178.64	1197.24	...	100.38	97
...
27/3/2022 8:50	0.49	97.16	1187.17	1242.59	1241.82	1201.37	1218.08	1220.33	1187.04	1206.99	...	101.43	97
27/3/2022 8:55	0.49	97.13	1186.94	1241.14	1240.23	1202.54	1213.46	1223.79	1184.49	1208.52	...	101.89	97
27/3/2022 9:00	0.49	97.11	1189.00	1246.05	1242.17	1204.69	1219.75	1226.00	1187.44	1209.16	...	101.91	97
27/3/2022 9:05	0.49	97.08	1190.00	1245.33	1244.11	1204.33	1220.12	1224.15	1190.11	1210.96	...	101.81	96
27/3/2022 15:50	0.45	97.69	1205.32	1256.40	1255.53	1221.99	1219.06	1244.19	1205.50	1224.44	...	111.56	106

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [6]: # Finding correlation data with speed and define as relational model. Otherwise define as individual model
RCS_A_corr = RCS_A.corr()[[speedtag]]
RCS_A_corr['sqrtn'] = RCS_A.corr(axis=0).apply(math.sqrt)
RCS_A_corr['t_score'] = RCS_A_corr[speedtag]*RCS_A_corr['sqrtn']
RCS_A_corr.columns = ['R_square', 'sqrtn', 't_score']
RCS_A_corr['t_score_clust'] = np.where(RCS_A_corr.t_score>3, 'relation', 'individual')
RCS_A_corr
```

```
Out[6]:
```

	R_square	sqrtn	t_score	t_score_clust
PI Tag				
TSO-DGS_A -4200-PT -736A -PV	0.215528	25.159491	5.422581	relation
TSO-ENGIN_A-4200-ST -438 -PV	0.940054	25.159491	23.651288	relation
TSO-ENGIN_A-4200-TE -851A -PV	0.865892	25.159491	21.785403	relation
TSO-ENGIN_A-4200-TE -852A -PV	0.890772	25.159491	22.411368	relation
TSO-ENGIN_A-4200-TE -853A -PV	0.872642	25.159491	21.955223	relation
...
TSO-ENGIN_A-4200-EGV-943A -FB	0.924522	25.159491	23.260512	relation
TSO-ENGIN_A-4200-MC -943A -PV	0.455387	25.159491	11.457313	relation
TSO-ENGIN_A-4200-EGV-943A -CMD	0.962589	25.159491	24.218251	relation
TSO-ENGIN_A-4200-OSM_BU	0.992929	25.159491	24.981594	relation
TSO-ENGIN_A-4200-PT -903A -PV	0.974998	25.159491	24.530459	relation

```
In [7]: # cleansing data that cannot calculate correlation with speed
RCS_A_corr = RCS_A_corr.dropna()
RCS_A_corr.info()

<class 'pandas.core.frame.DataFrame'>
Index: 103 entries, TSO-DGS_A -4200-PT -736A -PV to TSO-ENGIN_A-4200-PT -903A -PV
Data columns (total 4 columns):
# Column Non-Null Count Dtype
---
0 R_square 103 non-null float64
1 sqrtn 103 non-null float64
2 t_score 103 non-null float64
3 t_score_clust 103 non-null object
dtypes: float64(3), object(1)
memory usage: 4.0+ KB
```

Individual model

```
In [9]: # define individual model data frame for collect model parameter
RCS_A_individual_Model = pd.DataFrame()
RCS_A_individual_Model['Tag'] = RCS_A_corr[RCS_A_corr.t_score_clust=='individual'].index
```

```
In [10]: # define individual model data for model calculation
RCS_A_individual_data = RCS_A[RCS_A_individual_Model.Tag]
```

Check normal dist

```
In [11]: # transform data to normal distribution with box-cox transformation and JB testing
RCS_A_individual_data_boxcox = RCS_A_individual_data.copy()
alpha = 0.05

jbstest_list = list()
boxcox_jbstest_list = list()
boxcox_lambda_list = list()
for i in RCS_A_individual_Model.Tag:
    jbstest_list.append(stats.jarque_bera(RCS_A_individual_data[i]).pvalue)
    fitted_data, fitted_lambda = stats.boxcox(RCS_A_individual_data[i])
    RCS_A_individual_data_boxcox[i] = fitted_data
    boxcox_lambda_list.append(fitted_lambda)
    boxcox_jbstest_list.append(stats.jarque_bera(fitted_data).pvalue)
RCS_A_individual_Model['jb'] = jbstest_list
RCS_A_individual_Model['jb_cluster'] = np.where(RCS_A_individual_Model['jb']>alpha, 'normal dist', 'not normal dist')
RCS_A_individual_Model['boxcox_lambda'] = boxcox_lambda_list
RCS_A_individual_Model['boxcox_jb'] = boxcox_jbstest_list
RCS_A_individual_Model['boxcox_jb_cluster'] = np.where(RCS_A_individual_Model['boxcox_jb']>alpha, 'normal dist',
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

In [12]: RCS_A_individual_Model

	Tag	jb	jb_cluster	boxcox_lambda	boxcox_jb	boxcox_jb_cluster
0	TSO-ENGIN_A-4200-TV-351X-DIRECT	0.000000e+00	not normal dist	0.793654	0.000000e+00	not normal dist
1	TSO-ENGIN_A-4200-TV-352X-DIRECT	5.671628e-01	normal dist	0.218233	9.755243e-01	normal dist
2	TSO-ENGIN_A-4200-TV-353Y-DIRECT	0.000000e+00	not normal dist	-3.427713	3.927821e-02	not normal dist
3	TSO-LUBE_A-4200-TE-878A-PV	0.000000e+00	not normal dist	0.419358	0.000000e+00	not normal dist
4	TSO-LUBE_A-4200-TE-900A-PV	0.000000e+00	not normal dist	1.527228	0.000000e+00	not normal dist
5	TSO-ENGIN_A-4200-TV-354X-DIRECT	3.185556e-06	not normal dist	-1.523179	2.886459e-04	not normal dist
6	TSO-ENGIN_A-4200-TV-354Y-DIRECT	1.964229e-04	not normal dist	-4.543849	4.516012e-01	normal dist
7	TSO-ENGIN_A-4200-TV-355X-DIRECT	0.000000e+00	not normal dist	-5.296061	9.043044e-01	normal dist
8	TSO-DGS_A-4200-PDT-726A-PV	5.184957e-06	not normal dist	-0.540856	8.671032e-01	normal dist
9	TSO-DGS_A-4200-PDT-727A-PV	8.163342e-09	not normal dist	-0.119494	1.597616e-08	not normal dist
10	TSO-DGS_A-4200-PT-720A-PV	2.746285e-01	normal dist	-1.126759	9.782649e-01	normal dist
11	TSO-DGS_A-4200-PT-737A-PV	5.689497e-01	normal dist	1.015710	5.708580e-01	normal dist
12	TSO-ENCAN_A-4200-TE-903A-PV	2.199466e-10	not normal dist	-2.769376	9.263585e-01	normal dist
13	AP_HPC_ETA_Act	1.684049e-01	normal dist	1.663314	3.366953e-01	normal dist
14	AP_HPC_ETA_Nom	0.000000e+00	not normal dist	-30.985647	0.000000e+00	not normal dist
15	AP_HPC_Spd_Nom	0.000000e+00	not normal dist	-19.700843	NaN	not normal dist
16	TSO-COMU_A-4200-XT-752A-DIRECT	0.000000e+00	not normal dist	-8.817765	2.359460e-02	not normal dist
17	TSO-COMU_A-4200-YT-753A-DIRECT	3.888556e-12	not normal dist	-7.812008	9.861788e-01	normal dist
18	TSO-FUEL_A-4200-PT-900A-PV	0.000000e+00	not normal dist	84.222065	NaN	not normal dist
19	TSO-FUEL_A-4200-TT-900A-PV	0.000000e+00	not normal dist	-3.375578	2.093327e-01	normal dist
20	TSO-STATN-4200-FV-411-PV	0.000000e+00	not normal dist	713.934939	NaN	not normal dist

In [13]: RCS_A_individual_Model.groupby('boxcox_jb_cluster').agg({'jb':'count'}).rename(columns={'jb':'Total_Numbers'})

boxcox_jb_cluster	Total_Numbers
normal dist	10
not normal dist	20

In [14]: # pickup only normal distribution data for doing individual model
RCS_A_individual_Model = RCS_A_individual_Model[RCS_A_individual_Model.boxcox_jb_cluster=='normal dist']
RCS_A_individual_Model = RCS_A_individual_Model.dropna()
RCS_A_individual_Model

	Tag	jb	jb_cluster	boxcox_lambda	boxcox_jb	boxcox_jb_cluster
1	TSO-ENGIN_A-4200-TV-352X-DIRECT	5.671628e-01	normal dist	0.218233	0.975524	normal dist
6	TSO-ENGIN_A-4200-TV-354Y-DIRECT	1.964229e-04	not normal dist	-4.543849	0.451601	normal dist
7	TSO-ENGIN_A-4200-TV-355X-DIRECT	0.000000e+00	not normal dist	-5.296061	0.904304	normal dist
8	TSO-DGS_A-4200-PDT-726A-PV	5.184957e-06	not normal dist	-0.540856	0.867103	normal dist
10	TSO-DGS_A-4200-PT-720A-PV	2.746285e-01	normal dist	-1.126759	0.978265	normal dist
11	TSO-DGS_A-4200-PT-737A-PV	5.689497e-01	normal dist	1.015710	0.570858	normal dist
12	TSO-ENCAN_A-4200-TE-903A-PV	2.199466e-10	not normal dist	-2.769376	0.926359	normal dist
13	AP_HPC_ETA_Act	1.684049e-01	normal dist	1.663314	0.336695	normal dist
17	TSO-COMU_A-4200-YT-753A-DIRECT	3.888556e-12	not normal dist	-7.812008	0.986179	normal dist
19	TSO-FUEL_A-4200-TT-900A-PV	0.000000e+00	not normal dist	-3.375578	0.209333	normal dist

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Univariate k-sigma model

```
In [15]: #find parameter of univariate k-sigma mode consisting of mean and sd
RCS_A_individual_data_boxcox = RCS_A_individual_data_boxcox[RCS_A_individual_Model.Tag]
RCS_A_individual_Model['mean'] = RCS_A_individual_data_boxcox.mean().values
RCS_A_individual_Model['sd'] = RCS_A_individual_data_boxcox.std().values
RCS_A_individual_Model = RCS_A_individual_Model.dropna()
RCS_A_individual_Model = RCS_A_individual_Model.reset_index(drop=True)
RCS_A_individual_Model
```

```
Out[15]:
```

	Tag	jb	jb_cluster	boxcox_lambda	boxcox_jb	boxcox_jb_cluster	mean	sd
0	TSO-ENGIN_A-4200-TV -352X -DIRECT	5.671628e-01	normal dist	0.218233	0.975524	normal dist	-0.810868	3.700998e-02
1	TSO-ENGIN_A-4200-TV -354Y -DIRECT	1.964229e-04	not normal dist	-4.543849	0.451601	normal dist	-61.935508	6.623619e+00
2	TSO-ENGIN_A-4200-TV -355X -DIRECT	0.000000e+00	not normal dist	-5.296061	0.904304	normal dist	-10.551336	2.939421e+00
3	TSO-DGS_A -4200-PDT-726A -PV	5.184957e-06	not normal dist	-0.540856	0.867103	normal dist	-2.736408	2.200629e-01
4	TSO-DGS_A -4200-PT -720A -PV	2.746285e-01	normal dist	-1.126759	0.978265	normal dist	0.886888	1.684688e-05
5	TSO-DGS_A -4200-PT -737A -PV	5.689497e-01	normal dist	1.015710	0.570858	normal dist	-0.363502	5.492472e-02
6	TSO-ENCAN_A-4200-TE -903A -PV	2.199466e-10	not normal dist	-2.769376	0.926359	normal dist	0.361090	3.670068e-07
7	AP_HPC_ETA_Act	1.684049e-01	normal dist	1.663314	0.336695	normal dist	-0.462022	1.525119e-02
8	TSO-COMU_A -4200-YT -753A -DIRECT	3.888556e-12	not normal dist	-7.812008	0.986179	normal dist	-27.963370	5.549415e+00
9	TSO-FUEL_A -4200-TT -900A -PV	0.000000e+00	not normal dist	-3.375578	0.209333	normal dist	0.296239	9.621519e-07

isolation forest

```
In [16]: # define isolation forest model
isoforest = lambda X : IsolationForest(random_state=0,contamination='auto',n_estimators=10).fit(X.values.reshape(
```

```
In [17]: # train model
isoforest_list = list()
for i in RCS_A_individual_Model.Tag:
    isoforest_list.append(isoforest(RCS_A_individual_data_boxcox[i]))
RCS_A_individual_Model['isoforest_model'] = isoforest_list
```

```
In [18]: # view isolation forest model parameter
RCS_A_individual_Model.isoforest_model[0].get_params(deep=True)
```

```
Out[18]:
{'bootstrap': False,
 'contamination': 'auto',
 'max_features': 1.0,
 'max_samples': 'auto',
 'n_estimators': 10,
 'n_jobs': None,
 'random_state': 0,
 'verbose': 0,
 'warm_start': False}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [19]: # Check and summary individual model
RCS_A_individual_Model = RCS_A_individual_Model[RCS_A_individual_Model.boxcox_jb_cluster == 'normal dist']
RCS_A_individual_Model = RCS_A_individual_Model.set_index(['Tag'], drop=True)
RCS_A_individual_Model
```

```
Out[19]:
```

	jb	jb_cluster	boxcox_lambda	boxcox_jb	boxcox_jb_cluster	mean	sd	isoforest_mo
Tag								
TSO-ENGIN_A-4200-TV-352X-DIRECT	5.671628e-01	normal dist	0.218233	0.975524	normal dist	-0.810868	3.700998e-02	(ExtraTreeRegressor(max_depth max_features
TSO-ENGIN_A-4200-TV-354Y-DIRECT	1.964229e-04	not normal dist	-4.543849	0.451601	normal dist	-61.935508	6.623619e+00	(ExtraTreeRegressor(max_depth max_features
TSO-ENGIN_A-4200-TV-355X-DIRECT	0.000000e+00	not normal dist	-5.296061	0.904304	normal dist	-10.551336	2.939421e+00	(ExtraTreeRegressor(max_depth max_features
TSO-DGS_A-4200-PDT-726A-PV	5.184957e-06	not normal dist	-0.540856	0.867103	normal dist	-2.736408	2.200629e-01	(ExtraTreeRegressor(max_depth max_features
TSO-DGS_A-4200-PT-720A-PV	2.746285e-01	normal dist	-1.126759	0.978265	normal dist	0.886888	1.684688e-05	(ExtraTreeRegressor(max_depth max_features
TSO-DGS_A-4200-PT-737A-PV	5.689497e-01	normal dist	1.015710	0.570858	normal dist	-0.363502	5.492472e-02	(ExtraTreeRegressor(max_depth max_features
TSO-ENCAN_A-4200-TE-903A-PV	2.199466e-10	not normal dist	-2.769376	0.926359	normal dist	0.361090	3.670068e-07	(ExtraTreeRegressor(max_depth max_features
AP_HPC_ETA_Act	1.684049e-01	normal dist	1.663314	0.336695	normal dist	-0.462022	1.525119e-02	(ExtraTreeRegressor(max_depth max_features
TSO-COMU_A-4200-YT-753A-DIRECT	3.888556e-12	not normal dist	-7.812008	0.986179	normal dist	-27.963370	5.549415e+00	(ExtraTreeRegressor(max_depth max_features
TSO-FUEL_A-4200-TT-900A-PV	0.000000e+00	not normal dist	-3.375578	0.209333	normal dist	0.296239	9.621519e-07	(ExtraTreeRegressor(max_depth max_features

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

test performance - individual model

```

In [20]: ind_3sd_score_list = list()
ind_iso_score_list = list()
individual_label = RCS_A_individual_Model.index
for i in range(len(individual_label)):
    testtag = individual_label[i]
    X_sample = RCS_A_individual_data_boxcox[[testtag]].sample(100)
    X_sample['isnormal'] = 1

    scale = 1
    noise_pos = pd.DataFrame(np.random.normal(0, scale, int(len(X_sample)/2))+1.1*X_sample.max()[0], columns=[testtag, 'isnormal'], index=X_sample.index)
    noise_neg = pd.DataFrame(0.9*X_sample.min()[0]-np.random.normal(0, scale, int(len(X_sample)/2)), columns=[testtag, 'isnormal'], index=X_sample.index)

    X_noise = pd.concat([X_sample, noise_pos.sample(50), noise_neg.sample(50)], axis=0).sample(200).reset_index(drop=True)

    lb = RCS_A_individual_Model.loc[testtag]['mean']-3*RCS_A_individual_Model.loc[testtag]['sd']
    ub = RCS_A_individual_Model.loc[testtag]['mean']+3*RCS_A_individual_Model.loc[testtag]['sd']

    sd_result = X_noise.copy()
    sd_result['3sd'] = np.where((X_noise[testtag]<lb)|(X_noise[testtag]>ub), -1, 1)

    ind_3sd_score = {'F1': f1_score(sd_result['isnormal'], sd_result['3sd']),
                    'Precision': precision_score(sd_result['isnormal'], sd_result['3sd']),
                    'Recall': recall_score(sd_result['isnormal'], sd_result['3sd'])}
    ind_3sd_score_list.append(ind_3sd_score)

    iso_result = X_noise.copy()
    # iso_result['iso_score'] = RCS_A_individual_Model.loc[testtag]['isoforest_model'].score_samples(iso_result)
    iso_result['iso_score'] = RCS_A_individual_Model.loc[testtag]['isoforest_model'].decision_function(iso_result)
    iso_result['iso_prob'] = 1-(iso_result['iso_score']-iso_result['iso_score'].min())/(iso_result['iso_score'].max()-iso_result['iso_score'].min())
    iso_result['iso_prob_cluster'] = np.where(iso_result['iso_prob']>0.5, -1, 1)
    ind_iso_score = {'F1': f1_score(iso_result['isnormal'], iso_result['iso_prob_cluster']),
                    'Precision': precision_score(iso_result['isnormal'], iso_result['iso_prob_cluster']),
                    'Recall': recall_score(iso_result['isnormal'], iso_result['iso_prob_cluster'])}
    ind_iso_score_list.append(ind_iso_score)
ind_score_df = pd.concat([pd.DataFrame(ind_3sd_score_list), pd.DataFrame(ind_iso_score_list)], axis=1)
ind_score_df.index = individual_label
ind_score_df

```

```

Out[20]:

```

Tag	F1	Precision	Recall	F1	Precision	Recall
TSO-ENGIN_A-4200-TV-352X-DIRECT	0.970874	0.943396	1.00	0.921466	0.967033	0.88
TSO-ENGIN_A-4200-TV-354Y-DIRECT	0.666667	0.500000	1.00	0.761538	0.618750	0.99
TSO-ENGIN_A-4200-TV-355X-DIRECT	0.666667	0.500000	1.00	0.780000	0.780000	0.78
TSO-DGS_A-4200-PDT-726A-PV	0.800000	0.666667	1.00	0.830508	0.720588	0.98
TSO-DGS_A-4200-PT-720A-PV	1.000000	1.000000	1.00	0.936170	1.000000	0.88
TSO-DGS_A-4200-PT-737A-PV	0.925926	0.862069	1.00	0.942308	0.907407	0.98
TSO-ENCAN_A-4200-TE-903A-PV	1.000000	1.000000	1.00	0.974359	1.000000	0.95
AP_HPC_ETA_Act	0.980392	0.961538	1.00	0.943590	0.968421	0.92
TSO-COMU_A-4200-YT-753A-DIRECT	0.662207	0.497487	0.99	0.859729	0.785124	0.95
TSO-FUEL_A-4200-TT-900A-PV	1.000000	1.000000	1.00	0.913043	1.000000	0.84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Correlation model from hist

```
In [21]: # Define relational model data frame for collect parameter
RCS_A_relation_Model = pd.DataFrame()
RCS_A_relation_Model['Tag'] = RCS_A_corr[RCS_A_corr.t_score_clust=='relation'].index
RCS_A_relation_Model.loc[len(RCS_A_relation_Model.index)] = speedtag
RCS_A_relation_Model = RCS_A_relation_Model.drop_duplicates()
RCS_A_relation_Model
```

```
Out[21]:
```

	Tag
0	TSO-DGS_A -4200-PT -736A -PV
1	TSO-ENGIN_A-4200-ST -438 -PV
2	TSO-ENGIN_A-4200-TE -851A -PV
3	TSO-ENGIN_A-4200-TE -852A -PV
4	TSO-ENGIN_A-4200-TE -853A -PV
...	...
68	TSO-ENGIN_A-4200-EGV-943A -FB
69	TSO-ENGIN_A-4200-MC -943A -PV
70	TSO-ENGIN_A-4200-EGV-943A -CMD
71	TSO-ENGIN_A-4200-OSM_BU
72	TSO-ENGIN_A-4200-PT -903A -PV

70 rows × 1 columns

```
In [22]: # define relation model data for model calculation
RCS_A_relation_data = RCS_A[RCS_A_relation_Model.Tag]
RCS_A_relation_data = RCS_A_relation_data.T.drop_duplicates()
RCS_A_relation_data = RCS_A_relation_data.T
RCS_A_relation_data
```

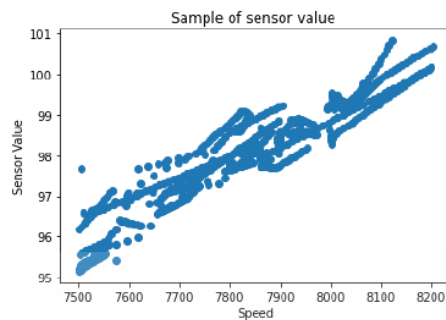
```
Out[22]:
```

	TSO-DGS_A -4200-PT -736A -PV	TSO-ENGIN_A-4200-ST -438 -PV	TSO-ENGIN_A-4200-TE -851A -PV	TSO-ENGIN_A-4200-TE -852A -PV	TSO-ENGIN_A-4200-TE -853A -PV	TSO-ENGIN_A-4200-TE -854A -PV	TSO-ENGIN_A-4200-TE -855A -PV	TSO-ENGIN_A-4200-TE -856A -PV	TSO-ENGIN_A-4200-TE -857A -PV	TSO-ENGIN_A-4200-TE -858A -PV	...	AP_Eng_T5_Pred_GF
23/3/2022 10:15	0.47	96.17	1176.35	1219.08	1223.05	1194.95	1193.72	1207.86	1173.05	1191.31	...	1192.58
23/3/2022 10:20	0.47	96.25	1178.54	1222.55	1225.16	1197.35	1195.38	1209.87	1174.26	1192.65	...	1191.64
23/3/2022 10:25	0.47	96.33	1178.27	1225.34	1225.60	1199.27	1197.09	1211.12	1173.18	1194.48	...	1195.16
23/3/2022 10:30	0.47	96.41	1180.19	1228.05	1232.08	1201.00	1198.84	1215.07	1176.66	1193.98	...	1199.53
23/3/2022 10:35	0.47	96.49	1181.30	1233.10	1232.13	1202.96	1201.54	1217.79	1178.64	1197.24	...	1203.60
...
27/3/2022 8:50	0.49	97.16	1187.17	1242.59	1241.82	1201.37	1218.08	1220.33	1187.04	1206.99	...	1208.60
27/3/2022 8:55	0.49	97.13	1186.94	1241.14	1240.23	1202.54	1213.46	1223.79	1184.49	1208.52	...	1210.67
27/3/2022 9:00	0.49	97.11	1189.00	1246.05	1242.17	1204.69	1219.75	1226.00	1187.44	1209.16	...	1211.01
27/3/2022 9:05	0.49	97.08	1190.00	1245.33	1244.11	1204.33	1220.12	1224.15	1190.11	1210.96	...	1211.59
27/3/2022 15:50	0.45	97.69	1205.32	1256.40	1255.53	1221.99	1219.06	1244.19	1205.50	1224.44	...	1233.27

633 rows × 70 columns

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
In [23]: plt.scatter(RCS_A_relation_data[speedtag],RCS_A_relation_data.iloc[:,1])
plt.xlabel('Speed')
plt.ylabel('Sensor Value')
plt.title('Sample of sensor value ')
plt.show()
```



```
In [24]: # Check regression assumption
rsquared_list = list()
db_list = list()
pearson_list = list()
resid_jb_list = list()
resid_wt_list = list()
DW_list = list()
for i in RCS_A_relation_Model.Tag:
    x = RCS_A_relation_data[i]
    model = sm.OLS(x,RCS_A_relation_data[speedtag])
    results = model.fit()
    rsquared_list.append(results.rsquared)
    pearson_list.append(RCS_A_relation_data[[i,speedtag]].corr().iloc[1][0])
    resid_wt_list.append(round(het_white(results.resid,sm.add_constant(x))[1,4]))
    db_list.append(durbin_watson(results.resid))
    resid_jb_list.append(stats.jarque_bera(results.resid).pvalue)
RCS_A_relation_Model['rsquared'] = rsquared_list
RCS_A_relation_Model['resid_db'] = db_list
RCS_A_relation_Model['resid_db_cluster'] = np.where(RCS_A_relation_Model.resid_db.between(1.5,2.5), 'no correlation', 'autocorr')
RCS_A_relation_Model['pearson'] = pearson_list
RCS_A_relation_Model['pearson_cluster'] = np.where(RCS_A_relation_Model.pearson>0.7, 'linearity', 'non linearity')
RCS_A_relation_Model['resid_jb'] = resid_jb_list
RCS_A_relation_Model['resid_jb_cluster'] = np.where(RCS_A_relation_Model.resid_jb>0.05, 'normal dist', 'not normal dist')
RCS_A_relation_Model['resid_wt'] = resid_wt_list
RCS_A_relation_Model['resid_wt_cluster'] = np.where(RCS_A_relation_Model.resid_wt>0.05, 'Homoscedasticity', 'Heteroscedasticity')
RCS_A_relation_Model
```

```
Out[24]:
```

	Tag	rsquared	resid_db	resid_db_cluster	pearson	pearson_cluster	resid_jb	resid_jb_cluster	resid_wt	resid_wt_cluster
0	TSO-DGS_A-4200-PT-736A-FV	0.998933	0.038059	autocorrelation	0.215528	non linearity	0.002068	not normal dist	0.0000	Heteroscedasticity
1	TSO-ENGIN_A-4200-ST-438-PV	0.998843	0.026782	autocorrelation	0.940054	linearity	0.000040	not normal dist	0.0000	Heteroscedasticity
2	TSO-ENGIN_A-4200-TE-851A-PV	0.999336	0.057102	autocorrelation	0.865892	linearity	0.000102	not normal dist	0.0000	Heteroscedasticity
3	TSO-ENGIN_A-4200-TE-852A-PV	0.999875	0.083557	autocorrelation	0.890772	linearity	0.000054	not normal dist	0.0000	Heteroscedasticity
4	TSO-ENGIN_A-4200-TE-853A-PV	0.998839	0.080682	autocorrelation	0.872642	linearity	0.001904	not normal dist	0.0000	Heteroscedasticity
...
66	TSO-ENGIN_A-4200-MC-943A-FV	0.915698	0.701503	autocorrelation	0.459387	non linearity	0.004373	not normal dist	0.0000	Heteroscedasticity
67	TSO-ENGIN_A-4200-EGV-943A-CMD	0.999778	0.024444	autocorrelation	0.962589	linearity	0.000002	not normal dist	0.0000	Heteroscedasticity
68	TSO-ENGIN_A-4200-OSM-BU	0.999992	0.351333	autocorrelation	0.992929	linearity	0.000077	not normal dist	0.1855	Homoscedasticity
69	TSO-ENGIN_A-4200-PT-903A-PV	0.999810	0.043877	autocorrelation	0.974998	linearity	0.000198	not normal dist	0.0000	Heteroscedasticity
70	AP_HPC_Spd_Act	1.000000	0.000000	autocorrelation	1.000000	linearity	NaN	not normal dist	1.0000	Homoscedasticity

71 rows x 10 columns

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Regression Control

```
In [28]: paramlist = list()
S_error_list = list()
SD_list = list()
S_error = lambda sd,rsq : sd*math.sqrt(1-rsq)
for i in RCS_A_relation_Model.index:
    param = stats.linregress(RCS_A_relation_data[speedtag],RCS_A_relation_data[i])
    paramlist.append([param.slope,param.intercept])
    predicted_val = param.slope*RCS_A_relation_data[speedtag]+param.intercept
    predicted_sd = predicted_val.std()
    SD_list.append(predicted_sd)
    S_error_list.append(S_error(predicted_sd,RCS_A_relation_Model.loc[i].rsquared))

dfS_error = pd.DataFrame()
dfS_error.index = RCS_A_relation_Model.index
dfS_error['param'] = paramlist
dfS_error['R_squared'] = RCS_A_relation_Model.rsquared
dfS_error['SD_pred'] = SD_list
dfS_error['S_error'] = S_error_list

dfS_error
```

```
Out[28]:
```

Tag	param	R_squared	SD_pred	S_error
TSO-ENGIN_A-4200-TE-854A -PV	[0.08517468890718616, 553.2766549522229]	0.999786	16.081416	0.235291
TSO-ENGIN_A-4200-TV -351Y -DIRECT	[0.00014664433537142163, -0.5932979741522332]	0.995119	0.027687	0.001934
TSO-FUEL_A -4200-FCV-964A -FB	[0.0067870861870350865, -4.779872348468679]	0.999934	1.281437	0.010392
TSO-ENGIN_A-4200-PDT-799 -PV	[0.00021447921043406254, -0.7157396580991676]	0.998872	0.040495	0.001360
AP_HPC_Head	[-2.1213167770876913, -9597.663100181584]	0.990536	400.515433	38.963895
AP_HPC_Spd_Act	[1.0, 0.0]	1.000000	188.805103	0.000000
TSO-COMU_A -4200-YT -751A -DIRECT	[6.59164489024452e-05, -0.0013747915158276447]	0.998991	0.012445	0.000395
AP_Eng_Fuel_Flow_Std_GF	[0.024815008250012937, -90.84456929256157]	0.999255	4.685200	0.127896
TSO-FUEL_A -4200-FCV-964A -CMD	[0.006787436617373773, -5.088618475699008]	0.999929	1.281503	0.010801

multivariate k-sigma

$$k_i = \sqrt{(\bar{X}_i - \bar{\pi})^T \Sigma^{-1} (\bar{X}_i - \bar{\pi})}$$

```
In [30]: covMatrixinv = lambda x,s : np.linalg.pinv(np.cov(np.stack([x,s],axis=0)))
covMatrixinv_list = list()
X_mean_list = list()
for i in RCS_A_relation_Model.index:
    X_mean_list.append(RCS_A_relation_data[i].mean())
    covMatrixinv_list.append(covMatrixinv(RCS_A_relation_data[i],RCS_A_relation_data[speedtag]))

dfmul_k = pd.DataFrame()
dfmul_k.index = RCS_A_relation_Model.index
dfmul_k['mean'] = X_mean_list
dfmul_k['invcov_speed'] = covMatrixinv_list
dfmul_k
```

```
Out[30]:
```

Tag	mean	invcov_speed
TSO-ENGIN_A-4200-TE-854A -PV	1219.975261	[[0.007081978153078225, -0.0006032052860359252...]]
TSO-ENGIN_A-4200-TV -351Y -DIRECT	0.554550	[[764.0070454334, -0.11203730539665036], [-0.1...]]
TSO-FUEL_A -4200-FCV-964A -FB	48.345545	[[7.102130784597069, -0.04820277374665596], [...]]
TSO-ENGIN_A-4200-PDT-799 -PV	0.963081	[[1327.946922508751, -0.2848170074383573], [-0...]]
AP_HPC_Head	7006.787773	[[2.396430064747749e-06, -5.083587301466741e-0...]]
AP_HPC_Spd_Act	7827.426367	[[7.013140695429386e-06, 7.013140695429384e-06...]]
TSO-COMU_A -4200-YT -751A -DIRECT	0.514581	[[3730.2541162466414, -0.24588510484705406], [...]]
AP_Eng_Fuel_Flow_Std_GF	103.393081	[[0.3126417644478856, -0.00775820796407292], [...]]
TSO-FUEL_A -4200-FCV-964A -CMD	48.039542	[[6.698907687397641, -0.04546841133385102], [...]]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Standard residual

```
In [32]: paramlist = list()
rmse_list = list()
sdres_list = list()
sdres_fn = lambda y,ypred : math.sqrt(((y-ypred)**2).sum())/(len(y)-2))
for i in RCS_A_relation_Model.index:
    param = stats.linregress(RCS_A_relation_data[speedtag],RCS_A_relation_data[i])
    paramlist.append([param.slope,param.intercept])
    predicted_val = param.slope*RCS_A_relation_data[speedtag]+param.intercept
    actual_val = RCS_A_relation_data[i]
    mse = mean_squared_error(actual_val, predicted_val)
    rmse_list.append(math.sqrt(mse))
    sdres = sdres_fn(actual_val, predicted_val)
    sdres_list.append(sdres)

dfsdres = pd.DataFrame()
dfsdres.index = RCS_A_relation_Model.index
dfsdres['param'] = paramlist
dfsdres['rmse'] = rmse_list
dfsdres['sdres'] = sdres_list
dfsdres
```

```
Out[32]:
```

	param	rmse	sdres
Tag			
TSO-ENGIN_A-4200-TE-854A -PV	[0.08517468890718616, 553.2766549522229]	11.873517	11.892319
TSO-ENGIN_A-4200-TV-351Y -DIRECT	[0.00014664433537142163, -0.5932979741522332]	0.036150	0.036207
TSO-FUEL_A-4200-FCV-964A -FB	[0.0067870861870350865, -4.779872348468679]	0.374940	0.375534
TSO-ENGIN_A-4200-PDT-799 -PV	[0.00021447921043406254, -0.7157396580991676]	0.027420	0.027463
AP_HPC_Head	[2.1213167770876913, -9597.663100181584]	645.467388	646.489506
AP_HPC_Spd_Act	[1.0, 0.0]	0.000000	0.000000
TSO-COMU_A-4200-YT-751A -DIRECT	[6.59164489024452e-05, -0.0013747915158276447]	0.016360	0.016386
AP_Eng_Fuel_Flow_Std_GF	[0.024815008250012937, -90.84456929256157]	1.787036	1.789865
TSO-FUEL_A-4200-FCV-964A -CMD	[0.006787436617373773, -5.088618475699008]	0.386060	0.386671

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

test performance - relation model

```
In [34]: dfs_error_score_list = list()
dfmul_k_score_list = list()
dfs_dres_score_list = list()
testtag = RCS_A_relation_Model[RCS_A_relation_Model.index!=speedtag].index

for tag in testtag:
    X_sample = RCS_A_relation_data[[tag,speedtag]].copy()
    X_sample['isnormal'] = 1

    scale = 5
    noise_pos = X_sample.copy()
    noise_pos[tag] = np.random.normal(0, scale, len(X_sample))+1.1*X_sample[tag].max()
    noise_pos['isnormal'] = -1

    noise_neg = X_sample.copy()
    noise_neg[tag] = 0.9*X_sample[tag].min()-np.random.normal(0, scale, len(X_sample))
    noise_neg['isnormal'] = -1

    X_noise = pd.concat([X_sample.sample(100),noise_pos.sample(50),noise_neg.sample(50)],axis=0).sample(200).re

    Z_fn = lambda y,ypred : (ypred-y)/dfs_error.loc[tag,'S_error']
    predicted_val = dfs_error.loc[tag,'param'][0]*X_noise[speedtag]+dfs_error.loc[tag,'param'][1]
    dfs_error_result = X_noise.copy()
    dfs_error_result['z_score'] = Z_fn(X_noise[tag],predicted_val)
    dfs_error_result['predict'] = predicted_val
    dfs_error_result['z_score_cluster'] = np.where(dfs_error_result['z_score'].abs()>10,-1,1)
    dfs_error_score = {'F1':f1_score(dfs_error_result['isnormal'], dfs_error_result['z_score_cluster']),\
    'Precision': precision_score(dfs_error_result['isnormal'], dfs_error_result['z_score_cluster']),\
    'Recall': recall_score(dfs_error_result['isnormal'], dfs_error_result['z_score_cluster'])}
    dfs_error_score_list.append(dfs_error_score)

    k_list = list()
    k_fn = lambda X_M : math.sqrt(np.matmul(np.matmul(X_M.T,dfmul_k.loc[speedtag,'invcov_speed']),X_M))
    x1 = (X_noise[tag]-dfmul_k.loc[tag,'mean']).values
    x2 = (X_noise[speedtag]-dfmul_k.loc[speedtag,'mean']).values
    X = np.stack((x1, x2), axis=1)
    K_i = list()
    for i in X:
        K_i.append(k_fn(i))
    dfmul_k_result = X_noise.copy()
    dfmul_k_result['k_score'] = K_i
    dfmul_k_result['k_score_cluster'] = np.where(dfmul_k_result['k_score'].abs()>0.43,-1,1)
    dfmul_k_score = {'F1':f1_score(dfmul_k_result['isnormal'], dfmul_k_result['k_score_cluster']),\
    'Precision': precision_score(dfmul_k_result['isnormal'], dfmul_k_result['k_score_cluster']),\
    'Recall': recall_score(dfmul_k_result['isnormal'], dfmul_k_result['k_score_cluster'])}
    dfmul_k_score_list.append(dfmul_k_score)

    Z_list = list()
    Z_fn = lambda y,ypred : (ypred-y)/dfs_dres.loc[tag,'sdres']
    predicted_val = dfs_dres.loc[tag,'param'][0]*X_noise[speedtag]+dfs_dres.loc[tag,'param'][1]
    Z_list.append(Z_fn(X_noise[tag],predicted_val))
    dfs_dres_result = X_noise.copy()
    dfs_dres_result['z_score'] = Z_list[0]
    dfs_dres_result['z_score_cluster'] = np.where(dfs_dres_result['z_score'].abs()>3,-1,1)
    dfs_dres_score = {'F1':f1_score(dfs_dres_result['isnormal'], dfs_dres_result['z_score_cluster']),\
    'Precision': precision_score(dfs_dres_result['isnormal'], dfs_dres_result['z_score_cluster']),\
    'Recall': recall_score(dfs_dres_result['isnormal'], dfs_dres_result['z_score_cluster'])}
    dfs_dres_score_list.append(dfs_dres_score)

relation_score_df = pd.concat([pd.DataFrame(dfs_error_score_list), pd.DataFrame(dfmul_k_score_list),pd.DataFrame(dfs_dres_score_list)])
relation_score_df.index = RCS_A_relation_Model.index[1:-1]
relation_score_df
```

```
Out[34]:
```

Tag	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
TSO-ENGIN_A-4200-TE-854A-PV	0.245614	1.000000	0.14	0.517073	0.504762	0.53	1.000000	1.000000	1.00
TSO-ENGIN_A-4200-TV-351Y-DIRECT	0.529412	1.000000	0.36	0.538462	0.518519	0.56	0.995025	0.990099	1.00
TSO-FUEL_A-4200-FCV-964A-FB	0.370968	0.958333	0.23	0.483092	0.467290	0.50	0.970874	0.943396	1.00
TSO-ENGIN_A-4200-PDT-799-PV	0.684211	1.000000	0.52	0.495146	0.481132	0.51	0.995025	0.990099	1.00
AP_HPC_Head	0.571429	1.000000	0.40	0.319328	1.000000	0.19	0.980392	0.961538	1.00
AP_HPC_Spd_Act	0.305085	1.000000	0.18	0.514563	0.500000	0.53	1.000000	1.000000	1.00
TSO-COMU_A-4200-YT-751A-DIRECT	0.709677	1.000000	0.55	0.461538	0.473684	0.45	1.000000	1.000000	1.00
AP_Eng_Fuel_Flow_Std_GF	0.384000	0.960000	0.24	0.478049	0.466667	0.49	0.970297	0.960784	0.98

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายณพล โรจนดำรงกุล
วัน เดือน ปีเกิด	29 เมษายน 2534
ที่อยู่ปัจจุบัน	59 ม.8 ถ.เสียงเมือง ต.นาป่า อ.เมือง จ.ชลบุรี
ประวัติการศึกษา	(2555) วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมไฟฟ้า เกรดเฉลี่ย 3.28 (มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน)
ทุนการศึกษาที่ได้รับ	ไม่มี
ผลงานทางวิชาการ	ไม่มี



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้