

การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐานอุตสาหกรรม  
ด้วยวิธีการเหมืองข้อมูล

**HARD DISK DRIVE QUALITY CHECKING BASED ON  
INDUSTRIAL STANDARD USING DATA MINING METHOD**

ชนดล สุชาติพงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า  
คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2557

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐานอุตสาหกรรม  
ด้วยวิธีการเหมืองข้อมูล

ชนดล สุชาติพงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2557

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

หัวข้อวิทยานิพนธ์

การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐาน  
อุตสาหกรรมด้วยวิธีการเหมืองข้อมูล  
Hard Disk Drive Quality Checking based on Industrial  
Standard Using Data Mining Method

ชื่อ-นามสกุล

นายชนดล สุชาติพงษ์

สาขาวิชา

วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษา

ผู้ช่วยศาสตราจารย์กฤษณ์ชนม์ ภูมิภิตติพิชญ์, Ph.D.

ปีการศึกษา

2557

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์บุญยัง ปลั่งกลาง, Dr.-Ing.)

.....กรรมการ  
(อาจารย์ณัฐภัทร พันธุ์คง, Ph.D.)

.....กรรมการ  
(อาจารย์พินิจ ศรีธร, Ph.D.)

.....กรรมการ  
(ผู้ช่วยศาสตราจารย์กฤษณ์ชนม์ ภูมิภิตติพิชญ์, Ph.D.)

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี อนุมัติวิทยานิพนธ์ฉบับนี้  
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์  
(ผู้ช่วยศาสตราจารย์ศิวกร อ่างทอง, Ph.D.)

วันที่ 6 เดือน มกราคม พ.ศ. 2558

หัวข้อวิทยานิพนธ์	การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐานอุตสาหกรรมด้วยวิธีการเหมืองข้อมูล
ชื่อ-นามสกุล	นายธนดล สุชาติพงศ์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์กฤษณ์ชนม์ ภูมิภักติพิชญ์, Ph.D.
ปีการศึกษา	2557

## บทคัดย่อ

ฮาร์ดดิสก์เป็นอุปกรณ์ที่ใช้เก็บข้อมูลที่มีความหลากหลาย เช่น เพลง ภาพยนตร์ หรือข้อมูลทางราชการ ผู้ใช้งานคาดหวังที่จะได้ฮาร์ดดิสก์ที่มีประสิทธิภาพสูง เนื่องจากจะถูกนำไปใช้ในการเก็บข้อมูลที่สำคัญ ในกระบวนการผลิตฮาร์ดดิสก์ที่มีความซับซ้อนอาจก่อให้เกิดความผิดพลาดในระหว่างการผลิตได้ จึงจำเป็นต้องมีการคัดแยกฮาร์ดดิสก์ที่ไม่ผ่านการทดสอบคุณภาพและความน่าเชื่อถือ เพื่อเป็นการควบคุมคุณภาพและเพิ่มความเชื่อมั่นสำหรับผู้ใช้งาน ในการทดสอบความน่าเชื่อถือจะมีการสุ่มตัวอย่างฮาร์ดดิสก์มาทดสอบ โดยจำลองลักษณะการใช้งานตามเงื่อนไขต่าง ๆ หากไม่ผ่านการทดสอบจะถูกนำมาวิเคราะห์หาสาเหตุการเสียโดยผู้เชี่ยวชาญ

วิทยานิพนธ์นี้ได้มีการทำเหมืองข้อมูลโดยใช้อัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID ในการเรียนรู้ประเภทการเสียของฮาร์ดดิสก์โดยใช้ข้อมูลที่ได้จากกระบวนการทดสอบคุณภาพมาทำการเลือกแอตทริบิวต์ที่สนใจและปรับแต่งข้อมูลให้เหมาะสม แล้วทำการเปรียบเทียบผลลัพธ์ที่ได้จากการทดสอบในแต่ละอัลกอริทึม เพื่อให้ได้แบบจำลองที่สามารถคัดแยกสาเหตุการเสียของฮาร์ดดิสก์ได้ถูกต้องมากที่สุด

ผลลัพธ์จากวิทยานิพนธ์พบว่าแบบจำลองที่ได้จากอัลกอริทึม C5.0 สามารถคัดแยกประเภทการเสียของฮาร์ดดิสก์ได้อย่างมีประสิทธิภาพ โดยมีค่าความถูกต้องคิดเป็นร้อยละ 99.79 และมีความสอดคล้องกับลำดับความสำคัญของแอตทริบิวต์ที่ถูกเลือก โดยผู้เชี่ยวชาญจึงทำให้แบบจำลองที่ได้มีความน่าเชื่อถือ แบบจำลองที่ได้นี้สามารถช่วยลดขั้นตอนทดสอบความน่าเชื่อถือและการหาสาเหตุการเสียของฮาร์ดดิสก์ได้ ผลลัพธ์ที่ได้สามารถนำไปปรับปรุงคุณภาพการผลิตได้รวดเร็วยิ่งขึ้น

**คำสำคัญ:** ฮาร์ดดิสก์ คุณภาพการผลิต การคัดแยกข้อมูล

<b>Thesis Title</b>	Hard Disk Drive Quality Checking based on Industrial Standard Using Data Mining Method
<b>Name-Surname</b>	Mr. Thanadon Suchatpong
<b>Program</b>	Electrical Engineering
<b>Thesis Advisor</b>	Assistant Professor Krischonme Bhumkittipich, Ph.D.
<b>Academic Year</b>	2014

## ABSTRACT

Hard disk drive (HDD) is data storage device for storage the music, movie or important information. The end-users expect to have the high performance of HDD in order to storage the important information. The complexity of production process was generated the error or problems during the production. It is necessary to check the HDD based on industrial standard and reliability in order to control the quality and reliability for users. HDDs were examined by the reliability test by using various samples under many operation conditions. If the result is failure, the HDD will be analyzed the cause of failure by experts.

This thesis focuses on the HDD failure mode classification using data mining method. It consists of C5.0, Neural Network, C & R Tree, SVM and CHAID algorithm. These algorithms are used to learn the failure mode using the attributes of the production process. Multiple learning techniques study the improvement on inducing a more accurate and sensitive model. The comparison of the each algorithm in order to get the best model that can classify the cause of failure.

The results show that the C5.0 algorithm is the best model for HDD failure mode classification with 99.79% accuracy. The priorities of attributes are consistent between C5.0 algorithm and expert persons that influence the resulting model's quality. This method can help to increase quality and reliability of production and to improve efficiency of the analysis process prior to deliver the products to the end-users.

**Keywords:** hard disk drive, quality and reliability, pattern classification

## กิตติกรรมประกาศ

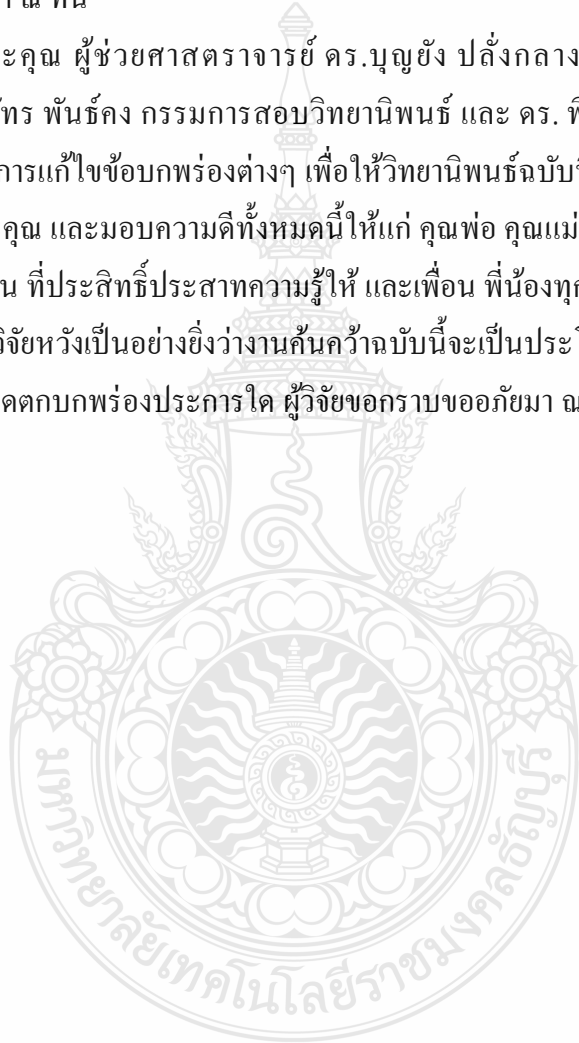
การศึกษาค้นคว้างานวิจัยฉบับนี้สำเร็จลุล่วงตามวัตถุประสงค์ที่ตั้งไว้ได้ด้วยความกรุณาของผู้ช่วยศาสตราจารย์ ดร.กฤษณ์ชนม์ ภูมิภักดีพิชญ์ อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำปรึกษา คำแนะนำ รวมถึงแนวทางในการปรับปรุงแก้ไขข้อบกพร่องต่างๆ จนสำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอกราบขอบพระคุณอย่างสูงมา ณ ที่นี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.บุญยัง ปลั่งกลาง ประธานกรรมการสอบวิทยานิพนธ์ ดร.ณัฐภัทร พันธุ์คง กรรมการสอบวิทยานิพนธ์ และ ดร. พิณิจ ศรีธร ผู้ทรงคุณวุฒิ ที่กรุณาให้คำแนะนำในการแก้ไขข้อบกพร่องต่างๆ เพื่อให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ และมอบความดีทั้งหมดนี้ให้แก่ คุณพ่อ คุณแม่ ที่คอยเลี้ยงดูอบรมสั่งสอน คณะครู-อาจารย์ทุกท่าน ที่ประสิทธิ์ประสาทความรู้ให้ และเพื่อน พี่น้องทุกคนที่ทำให้กำลังใจเสมอมา

สุดท้ายนี้ ผู้วิจัยหวังเป็นอย่างยิ่งว่างานค้นคว้าฉบับนี้จะเป็นประโยชน์สำหรับผู้ที่เกี่ยวข้อง หากการค้นคว้าในครั้งนี้ขาดตกบกพร่องประการใด ผู้วิจัยขอกราบขอภัยมา ณ โอกาสนี้ด้วย

ธนดล สุชาติพงษ์



## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	(3)
บทคัดย่อภาษาอังกฤษ.....	(4)
กิตติกรรมประกาศ.....	(5)
สารบัญ.....	(6)
สารบัญตาราง.....	(8)
สารบัญรูป.....	(9)
บทที่ 1 บทนำ.....	12
1.1 ความเป็นมาและความสำคัญของปัญหา .....	12
1.2 วัตถุประสงค์ของการวิจัย.....	13
1.3 สมมติฐานของงานวิจัย .....	13
1.4 ขอบเขตของการวิจัย .....	13
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	14
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	15
2.1 ส่วนประกอบของฮาร์ดดิสก์.....	15
2.2 มาตรฐานการทดสอบในอุตสาหกรรมฮาร์ดดิสก์ .....	19
2.3 การวิเคราะห์อาการเสียของฮาร์ดดิสก์ (Failure Mode Analysis) .....	22
2.4 การทำเหมืองข้อมูล (Data Mining).....	25
2.5 เทคนิคการเพิ่มประสิทธิภาพการคัดแยกข้อมูล.....	45
2.6 เทคนิคการวัดประสิทธิภาพ .....	47
2.7 การสำรวจเอกสาร (Literature Reviews) .....	49
บทที่ 3 วิธีดำเนินการวิจัย .....	53
3.1 ขั้นตอนการดำเนินงานวิจัย .....	53
3.2 การเตรียมข้อมูล.....	53
3.3 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม .....	58
3.4 ออกแบบการทดลอง.....	61
3.5 สรุปวิธีดำเนินงานวิจัย .....	63

## สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการทดลอง .....	64
4.1 ผลการสร้างแบบจำลอง .....	64
4.2 การทดลองหาประสิทธิภาพของแบบจำลอง .....	77
4.3 สรุปผลการทดลอง.....	84
บทที่ 5 สรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะ .....	86
5.1 สรุปผลการวิจัยและการอภิปรายผล.....	86
5.2 ข้อเสนอแนะ .....	87
5.3 ปัญหาและอุปสรรค .....	87
รายการอ้างอิง.....	88
ภาคผนวก.....	91
ภาคผนวก ก ผลการจัดเตรียมข้อมูล.....	92
ภาคผนวก ข คู่มือการใช้งาน โปรแกรม Clementine 12.0 .....	106
ภาคผนวก ค ผลงานตีพิมพ์เผยแพร่.....	115
ประวัติผู้เขียน.....	142



## สารบัญตาราง

	หน้า
ตารางที่ 2.1 มาตรฐานที่สำคัญในกระบวนการทดสอบความน่าเชื่อถือของฮาร์ดดิสก์ .....	22
ตารางที่ 2.2 ตัวอย่างชุดข้อมูลประกอบการตัดสินใจเล่นกีฬาชนิดหนึ่ง .....	34
ตารางที่ 2.3 ตัวอย่างชุดข้อมูลการซื้อเครื่องปรับอากาศ .....	43
ตารางที่ 2.4 ตัวอย่างระบบการแบ่งแยกแบบ Confusion Matrix .....	49
ตารางที่ 3.1 ประเภทการเสียทั้งหมดที่พบย้อนหลังเป็นเวลา 1 ปี .....	54
ตารางที่ 3.2 รายละเอียดการแปรค่าแอดทริบิวต์ที่นำมาทดลอง .....	55
ตารางที่ 3.3 การจัดกลุ่มความสำคัญของแอดทริบิวต์โดยผู้เชี่ยวชาญ .....	56
ตารางที่ 3.4 จำนวนชุดข้อมูลสำหรับสอน และชุดข้อมูลสำหรับทดสอบ .....	62
ตารางที่ 4.1 แสดงแอดทริบิวต์ที่มีการใช้งานในกระบวนการสร้างแบบจำลอง C5.0 .....	70
ตารางที่ 4.2 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของ แบบจำลอง C5.0 .....	78
ตารางที่ 4.3 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของ แบบจำลอง Neural Network .....	79
ตารางที่ 4.4 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของ แบบจำลอง C&R Tree .....	81
ตารางที่ 4.5 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของ แบบจำลอง SVM .....	82
ตารางที่ 4.6 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของ แบบจำลอง CHAID .....	84
ตารางที่ 4.7 ค่าความถูกต้องโดยรวมของแต่ละแบบจำลอง .....	85

## สารบัญรูป

	หน้า
รูปที่ 2.1 ส่วนประกอบของฮาร์ดดิสก์.....	15
รูปที่ 2.2 การอ่าน/เขียนข้อมูลลงบนจานแม่เหล็ก .....	16
รูปที่ 2.3 การจัดแบ่งพื้นที่บนแผ่นจานแม่เหล็กสำหรับบันทึกข้อมูล .....	17
รูปที่ 2.4 การจัดแบ่งพื้นที่บนแผ่นจานแม่เหล็กสำหรับบันทึกข้อมูลแบบแบ่งโซน.....	17
รูปที่ 2.5 การทำงานของแขนหัวอ่าน .....	18
รูปที่ 2.6 กระบวนการทดสอบคุณภาพฮาร์ดดิสก์ที่ใช้งานในปัจจุบัน.....	19
รูปที่ 2.7 กระบวนการทำงานของเหมืองข้อมูล .....	25
รูปที่ 2.8 ขั้นตอนการสร้างโมเดล การทดสอบโมเดล และการใช้โมเดลเป็นตัวจำแนกข้อมูล .....	31
รูปที่ 2.9 ส่วนประกอบของต้นไม้ตัดสินใจ .....	32
รูปที่ 2.10 ชุดข้อมูลของ play .....	35
รูปที่ 2.11 ชุดข้อมูลที่นำมาใช้คำนวณของ outlook.....	35
รูปที่ 2.12 ชุดข้อมูลของ outlook.....	36
รูปที่ 2.13 ชุดข้อมูลของ temperature.....	37
รูปที่ 2.14 ชุดข้อมูลของ humidity.....	37
รูปที่ 2.15 ชุดข้อมูลของ windy.....	38
รูปที่ 2.16 ลักษณะการแยกโหนดในระดับที่ 2 ต่อจาก outlook (root node) .....	38
รูปที่ 2.17 ลักษณะการแยกโหนดในระดับที่ 3 (temperature) ต่อจาก sunny .....	39
รูปที่ 2.18 ลักษณะการแยกโหนดในระดับที่ 3 (windy) ต่อจาก sunny .....	40
รูปที่ 2.19 ลักษณะการแยกโหนดในระดับที่ 3 (humidity) ต่อจาก sunny.....	40
รูปที่ 2.20 ลักษณะของต้นไม้ตัดสินใจที่ได้จากข้อมูลการเล่นกีฬา .....	41
รูปที่ 2.21 ตัวอย่างการแบ่งกลุ่มข้อมูลที่ Attribute Age.....	44
รูปที่ 2.22 เทคนิค Bagging เพื่อเพิ่มประสิทธิภาพการจำแนกข้อมูล .....	46
รูปที่ 2.23 เทคนิค Boosting เพื่อเพิ่มประสิทธิภาพการจำแนกข้อมูล .....	46
รูปที่ 2.24 การเลือกสุ่มข้อมูลแบบร้อยละ .....	47
รูปที่ 2.25 การเลือกสุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม เมื่อ K = 4 .....	48
รูปที่ 2.26 การเลือกสุ่มข้อมูลแบบ Leave-one-out Cross Validation .....	48

## สารบัญรูป (ต่อ)

หน้า

รูปที่ 2.27 การตรวจสอบเบื้องต้นในกระบวนการผลิต.....	50
รูปที่ 2.28 แบบจำลองที่ใช้โครงข่ายประสาทเทียมแบบเดียวในการทำนาย.....	51
รูปที่ 2.29 แบบจำลองโมเดลของ คณะกรรมการ Network.....	51
รูปที่ 2.30 กระบวนการวิเคราะห์สาเหตุการเสียชีวิตสู่กระบวนการปรับปรุงคุณภาพการผลิต .....	52
รูปที่ 3.1 โครงสร้างข้อมูลที่จะใช้กับ โปรแกรมประมวลผล .....	57
รูปที่ 3.2 หน้าหลักของโปรแกรม Clementine 12.0 .....	58
รูปที่ 3.3 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม C5.0.....	59
รูปที่ 3.4 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม Neural Network.....	59
รูปที่ 3.5 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม C&R Tree.....	60
รูปที่ 3.6 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม SVM.....	60
รูปที่ 3.7 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม CHAID.....	61
รูปที่ 3.8 กระบวนการสร้างแบบจำลองชนิดต่างๆ .....	62
รูปที่ 3.9 ตัวอย่างการหาค่าความถูกต้องของแบบจำลอง .....	63
รูปที่ 4.1 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง C5.0.....	64
รูปที่ 4.2 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 1 .....	65
รูปที่ 4.3 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 2 .....	65
รูปที่ 4.4 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 3 .....	66
รูปที่ 4.5 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 4.....	66
รูปที่ 4.6 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 5 .....	67
รูปที่ 4.7 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 6.....	67
รูปที่ 4.8 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 7 .....	68
รูปที่ 4.9 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 8 .....	68
รูปที่ 4.10 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 9 .....	69
รูปที่ 4.11 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 10 .....	69
รูปที่ 4.12 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง Neural Network.....	74
รูปที่ 4.13 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง C&R Tree.....	74

## สารบัญรูป (ต่อ)

	หน้า
รูปที่ 4.14 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของแบบจำลอง C&R Tree .....	75
รูปที่ 4.15 ลำดับความสำคัญของแอตทริบิวท์ที่นำมาประมวลผลแบบจำลอง SVM.....	75
รูปที่ 4.16 ลำดับความสำคัญของแอตทริบิวท์ที่นำมาประมวลผลแบบจำลอง CHAID.....	76
รูปที่ 4.17 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของแบบจำลอง CHAID .....	76
รูปที่ 4.18 ผลการทดสอบของแบบจำลอง C5.0.....	77
รูปที่ 4.19 ผลการทดสอบของแบบจำลอง Neural Network.....	79
รูปที่ 4.20 ผลการทดสอบของแบบจำลอง C&R Tree.....	80
รูปที่ 4.21 ผลการทดสอบของแบบจำลอง SVM.....	82
รูปที่ 4.22 ผลการทดสอบของแบบจำลอง CHAID .....	83



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ความก้าวหน้าทางด้านเทคโนโลยีสารสนเทศทำให้เกิดความต้องการฮาร์ดดิสก์เพื่อจัดเก็บข้อมูลเป็นจำนวนมาก โดยในปัจจุบันข้อมูลที่ถูกจัดเก็บในฮาร์ดดิสก์มีความสำคัญมาก เช่น ข้อมูลทางการค้าของบริษัทหรือห้างร้านต่าง ๆ ข้อมูลการเงินส่วนบุคคล และข้อมูลทางราชการ เป็นต้น ซึ่งแสดงให้เห็นว่าฮาร์ดดิสก์เป็นอุปกรณ์ที่มีความจำเป็นมาก จึงทำให้ผู้ใช้งานมีความต้องการฮาร์ดดิสก์ที่มีประสิทธิภาพและมีคุณภาพสูง และใช้งานได้อย่างคุ้มค่าคุ้มราคา เหตุผลดังกล่าวทำให้อุตสาหกรรมการผลิตฮาร์ดดิสก์ต้องมีการพัฒนาเพิ่มขึ้นทั้งในด้านความจุ กำลังการผลิต คุณภาพในกระบวนการผลิต และคุณภาพของสินค้า ดังนั้นขั้นตอนการผลิตฮาร์ดดิสก์ในปัจจุบันจะต้องผ่านกระบวนการการผลิตที่มีความละเอียดอ่อนมากยิ่งขึ้น เนื่องจากอุปกรณ์ภายในมีขนาดเล็กลงเพื่อเพิ่มความจุในการเก็บข้อมูล จึงต้องมีการทดสอบคุณภาพที่ซับซ้อนในหัวข้อต่าง ๆ ตามมาตรฐานกระบวนการทดสอบตามขั้นตอน (Process Test) และหลังจากนั้นจะถูกสุ่มมาทดสอบคุณภาพอีกครั้งหนึ่งซึ่งเรียกว่า การทดสอบความน่าเชื่อถือ (Reliability Test) ก่อนส่งมอบให้ผู้ใช้งานต่อไป การทดสอบดังกล่าวมีความสำคัญมากเนื่องจากยังพบว่ายังมีฮาร์ดดิสก์ที่ไม่ผ่านการทดสอบ ถึงแม้ฮาร์ดดิสก์ดังกล่าวจะผ่านกระบวนการทดสอบมาแล้วซึ่งหากขาดการทดสอบความน่าเชื่อถือ อาจส่งผลให้ฮาร์ดดิสก์ที่ไม่มีคุณภาพถูกส่งมอบให้ผู้ใช้งาน แต่การทำการทดสอบความน่าเชื่อถืออาจใช้เวลานานและเครื่องทดสอบดังกล่าวก็มีราคาสูง อีกทั้งสามารถสุ่มมาทดสอบได้เพียง 10-15% และยังคงจับฮาร์ดดิสก์ที่มีโอกาสเสียได้อีก 0.73% จากกลุ่มที่เข้าทดสอบ [1] ในการทดสอบความน่าเชื่อถือ เพื่อแก้ปัญหาให้ตรงจุด จำเป็นต้องอาศัยข้อมูลที่บ่งชี้ถึงอาการเสียของฮาร์ดดิสก์ ซึ่งได้มีการรวบรวมไว้ในขั้นตอนการตรวจสอบคุณภาพ

การวิเคราะห์อาการเสียของฮาร์ดดิสก์ จำเป็นต้องใช้ผู้เชี่ยวชาญวิเคราะห์ข้อมูลต่าง ๆ จากกระบวนการผลิตและทดสอบ ซึ่งผู้เชี่ยวชาญไม่สามารถวิเคราะห์ฮาร์ดดิสก์ได้ทั้งหมด เนื่องจากข้อมูลที่ใช้ในการวิเคราะห์มีจำนวนมาก ซึ่งการวิเคราะห์ฮาร์ดดิสก์แต่ละตัวที่ไม่ผ่านการทดสอบความน่าเชื่อถือ ต้องใช้เวลานานในการวิเคราะห์ และโดยเฉพาะอย่างยิ่งการวิเคราะห์โดยอาศัยผู้เชี่ยวชาญนั้นอาจประสบปัญหาของข้อจำกัดทางด้านองค์ความรู้ (Knowledge Based) [2] ผู้เชี่ยวชาญอาจมีความรู้อยู่ในขอบเขตหนึ่ง ซึ่งอาจจะไม่เพียงพอต่อการวิเคราะห์ลักษณะสาเหตุการเสียใหม่ ๆ ที่เกิดขึ้น

งานวิจัยนี้ตระหนักถึงความสำคัญของคุณภาพของฮาร์ดดิสก์ทุกตัวที่จะส่งไปถึงผู้ใช้งาน สามารถใช้งานได้อย่างเต็มประสิทธิภาพ และสามารถแยกประเภทการเสียบของฮาร์ดดิสก์เพื่อหาวิธี ป้องกันเหตุการณ์เสียเหล่านั้นและปรับปรุงคุณภาพการผลิตได้เร็วขึ้น จึงทำการศึกษาวิเคราะห์ข้อมูล ที่มีการบันทึกไว้ในกระบวนการผลิตของฮาร์ดดิสก์แต่ละตัวมาจัดกลุ่ม ปรับแต่งข้อมูล เพื่อเสนอ วิธีการที่สามารถคัดกรองฮาร์ดดิสก์ที่มีโอกาสเสียให้ได้มากที่สุด โดยใช้อัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID ในการเรียนรู้ประเภทการเสียบของฮาร์ดดิสก์ มาช่วยในการทำหน้าที่ทดสอบความน่าเชื่อถือ โดยให้อัลกอริทึมต่าง ๆ หาความสัมพันธ์ระหว่างค่าที่ได้จากการ ทดสอบตามขั้นตอน กับผลการทดสอบความน่าเชื่อถือ ทำให้สามารถนำค่าจากการทดสอบตาม ขั้นตอนของฮาร์ดดิสก์ใด ๆ มาป้อนเข้าแบบจำลองที่ได้จากอัลกอริทึมและทำนายว่าฮาร์ดดิสก์ตัว ดังกล่าวจะผ่านการทดสอบความน่าเชื่อถือหรือไม่ และมีลักษณะการเสียบอย่างไร ซึ่งสามารถลด ขั้นตอนการทดสอบความน่าเชื่อถือ และขั้นตอนการวิเคราะห์หาประเภทการเสียบของฮาร์ดดิสก์ได้ การใช้แบบจำลองดังกล่าวจึง สามารถนำฮาร์ดดิสก์ทุกตัวผ่านการทดสอบความน่าเชื่อถือ โดยไม่ จำเป็นต้องสุ่มทดสอบ ซึ่งช่วยให้มั่นใจได้ว่าฮาร์ดดิสก์ที่ไม่ได้คุณภาพจะไม่ถูกส่งมอบให้ผู้ใช้งาน

## 1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อหาตัวแปรและสร้างแบบจำลองในการแยกประเภทอาการเสียบของฮาร์ดดิสก์จาก ข้อมูลที่ใช้ในการทดสอบคุณภาพ

1.2.2 เพื่อเพิ่มประสิทธิภาพและลดระยะเวลาของกระบวนการวิเคราะห์อาการเสียบของ ฮาร์ดดิสก์ก่อนส่งมอบให้ผู้ใช้งาน

## 1.3 สมมติฐานของงานวิจัย

การใช้แบบจำลองต้นไม้ตัดสินใจชนิด C5.0 สามารถแยกประเภทการเสียบของฮาร์ดดิสก์ได้ ดีกว่าแบบจำลองอื่น ๆ

## 1.4 ขอบเขตของการวิจัย

1.4.1 วิเคราะห์ข้อมูลจากการตรวจสอบคุณภาพตามมาตรฐานเพื่อสรุปสาเหตุการเสียบของ ฮาร์ดดิสก์ ในส่วนของ Enterprise Storage ขนาด 3.5 นิ้ว จำนวน 1 ผลิตภัณฑ์ จากข้อมูลที่บันทึกไว้ ย้อนหลัง 1 ปี

1.4.2 จำแนกประเภทการเสียบของฮาร์ดดิสก์โดยได้ความถูกต้องมากกว่า 90%

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้แบบจำลองที่สามารถจำแนกประเภทการเสียชีวิตของฮาร์ดดิสก์ได้ และช่วยลดระยะเวลาในการวิเคราะห์สาเหตุการเสียชีวิตของฮาร์ดดิสก์ที่มีจำนวนมาก

1.5.2 นำผลลัพธ์ที่ได้จากแบบจำลองไปปรับปรุงคุณภาพการผลิตได้รวดเร็วยิ่งขึ้น

1.5.3 สามารถนำแบบจำลองไปพัฒนาเป็นซอฟต์แวร์เพื่อเพิ่มแนวทางการป้องกันไม่ให้ฮาร์ดดิสก์ที่ไม่ได้คุณภาพถูกส่งมอบให้กับผู้ใช้งาน



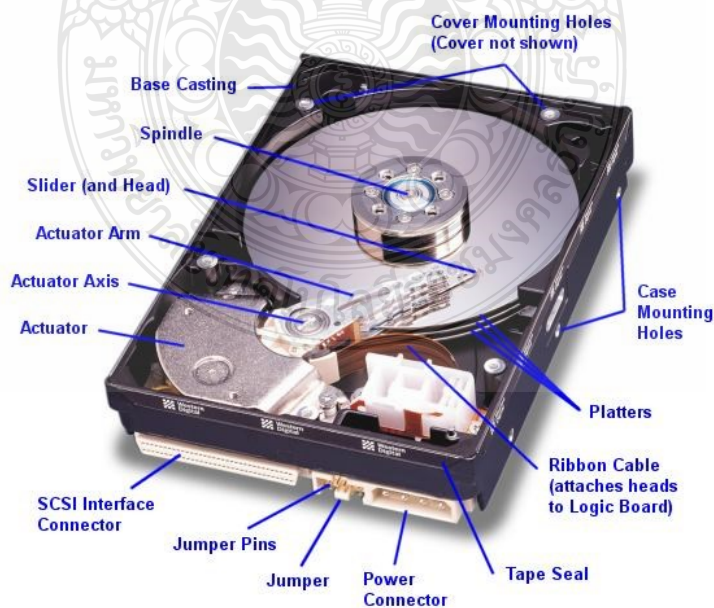
## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยจะกล่าวถึงส่วนประกอบของฮาร์ดดิสก์ การตรวจสอบคุณภาพภายในอุตสาหกรรมฮาร์ดดิสก์ การนำข้อมูลมาใช้ และเทคนิคที่ใช้ในการคัดแยกประเภทการเสียดัวยอัลกอริทึม C5.0 รวมถึงเทคนิคอื่น ๆ ที่ได้กล่าวถึงในงานวิจัย โดยมีรายละเอียดดังนี้

#### 2.1 ส่วนประกอบของฮาร์ดดิสก์

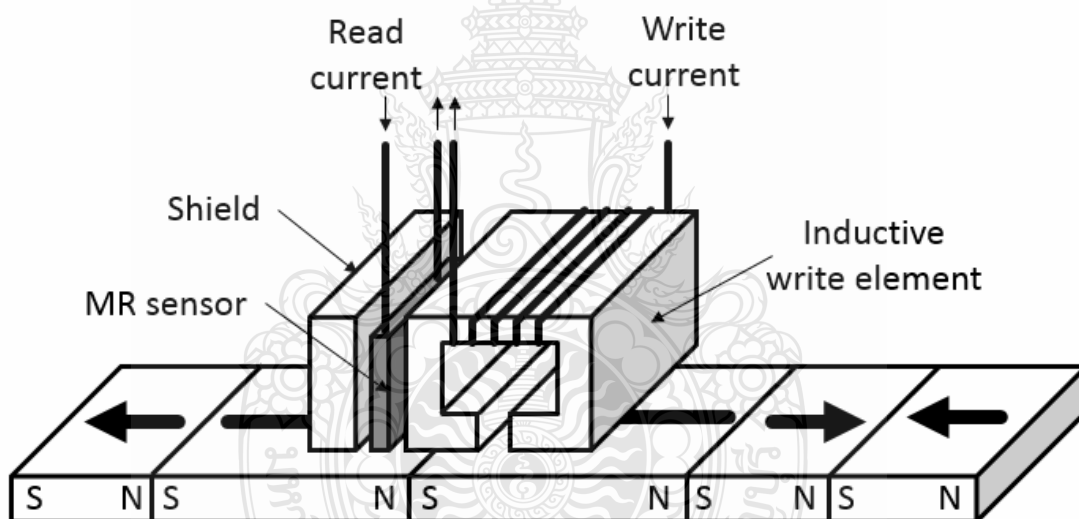
ส่วนประกอบของฮาร์ดดิสก์ [3] สามารถแบ่งออกเป็น 4 ส่วนที่สำคัญ ดังต่อไปนี้ ส่วนประกอบทางแม่เหล็ก (magnetic components) ซึ่งส่วนเครื่องจักรกล (mechanical components) ส่วนประกอบกลไฟฟ้า (electro-mechanical components) และ อุปกรณ์อิเล็กทรอนิกส์ (electronic device) ซึ่งส่วนประกอบทางแม่เหล็ก ได้แก่ แผ่นจานแม่เหล็ก และหัวอ่าน/เขียน ข้อมูล เป็นองค์ประกอบหลักที่ใช้ในการจัดเก็บและอ่านข้อมูล ซึ่งในฮาร์ดดิสก์จะมีการใช้หัวอ่าน/เขียน ในจัดเก็บข้อมูลบนพื้นผิวของจานที่เคลือบด้วยสารแม่เหล็ก และยังมีอุปกรณ์ที่สำคัญอีกหลายอย่าง เช่น มอเตอร์สำหรับหมุนจานแม่เหล็ก แขนหัวอ่านสำหรับนำหัวอ่าน/เขียน ไปยังตำแหน่งที่ต้องการ รูปที่ 2.1 แสดงให้เห็นถึงองค์ประกอบที่สำคัญที่พบในฮาร์ดดิสก์ทั่วไป



รูปที่ 2.1 ส่วนประกอบของฮาร์ดดิสก์ [4]



2.1.1 หัวอ่าน (Read/Write Head) เป็นส่วนที่ใช้ในการอ่าน/เขียนข้อมูล มีขนาดเล็กและมีความซับซ้อน ภายในหัวอ่านมีลักษณะเป็นขดลวด โดยในการอ่าน/เขียนข้อมูล จะมีการป้อนแรงดันไฟฟ้าเข้าสู่ขดลวดทำให้เกิดแรงเหนี่ยวนำไปเปลี่ยนโครงสร้างของสารแม่เหล็กที่ฉาบบนแผ่นจานแม่เหล็ก ให้มีรูปแบบของการบันทึกข้อมูล ซึ่งเทคโนโลยีของการอ่าน/เขียนนี้เรียกว่า MR (Magnical Recorder) ก่อนจะมาเป็น MR เทคโนโลยีที่เกิดขึ้นก่อนหน้านี้ คือ หัวอ่านแบบ Thin Film และหลังจาก MR แล้ว giant magneto-resistance (GMR) ก็ถูกพัฒนาตามมา จำนวนหัวอ่านในฮาร์ดดิสก์ตัวหนึ่งจะขึ้นอยู่กับจำนวนของแผ่นจานแม่เหล็ก โดยหนึ่งด้านของแผ่นจากแม่เหล็กจะต้องมีหัวอ่านจำนวน 1 หัวเสมอ ฮาร์ดดิสก์โดยส่วนใหญ่ที่มีแผ่นจานแม่เหล็กตั้งแต่ 1-4 แผ่น จะมีจำนวนหัวอ่านตั้งแต่ 2-8 หัว การอ่านและเขียนข้อมูลลงบนจานแม่เหล็กอาศัยหลักการตามทฤษฎีแม่เหล็กไฟฟ้า ซึ่งทุกครั้งที่มีการอ่านหรือเขียนข้อมูลจะมีเพียงหนึ่งหัวอ่านเท่านั้นที่มีการเข้าถึงข้อมูลได้ ซึ่งวิธีการอ่านข้อมูลและบันทึกข้อมูลลงบนแผ่นจานแม่เหล็กดังแสดงในรูปที่ 2.2

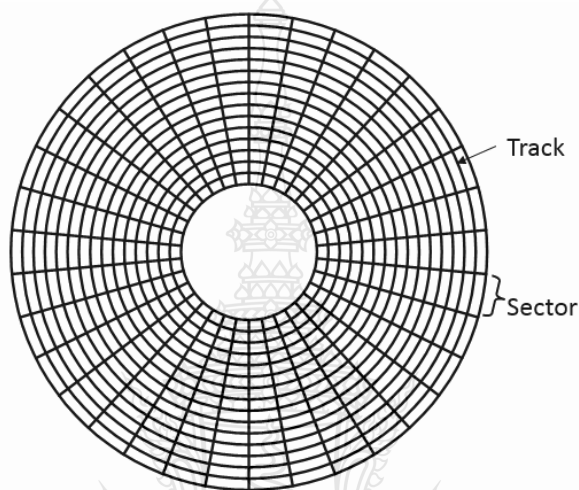


รูปที่ 2.2 การอ่าน/เขียนข้อมูลลงบนจานแม่เหล็ก [3]

2.1.1.1 การเขียนข้อมูลทุกครั้งที่หัวอ่านทำการเขียนข้อมูลจะมีกระแสไฟฟ้าป้อนให้กับส่วนของขดลวดที่ปลายหัวเขียน ทำให้ถูกเปลี่ยนสถานะเป็นแม่เหล็กไฟฟ้าชั่วคราวและจะไปเหนี่ยวนำให้สารแม่เหล็กบนจานแม่เหล็กมีการจัดเรียงทิศทางตามรูปแบบข้อมูลที่ต้องการบันทึก ที่มีสถานะเป็น 0 กับ 1

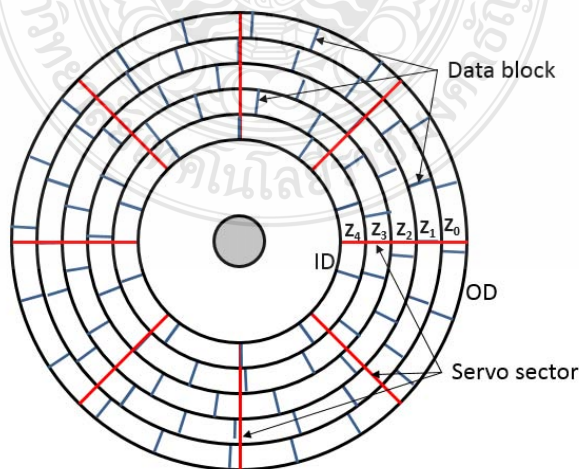
2.1.1.2 การอ่านข้อมูลของหัวอ่านจะเป็นไปในทางกลับกัน คือ หัวอ่านจะอยู่ในสถานะที่ไม่มีกระแสไฟฟ้า จึงเกิดการเหนี่ยวนำจากกระแสไฟฟ้าขึ้นที่ขดลวด ซึ่งค่ากระแสที่อ่านได้คือ ข้อมูลที่อ่านได้จากจานแม่เหล็ก

2.1.2 แผ่นจานแม่เหล็ก (Platters) เป็นพื้นที่ที่ใช้ในการเก็บข้อมูลอย่างเป็นระเบียบ โดยเนื้อที่บนแผ่นจานแม่เหล็กมีลักษณะกลมแบนจะถูกแบ่งออกเป็นแถบ หรือแทรค (Track) บนที่กข้อมูลรูปวงแหวนที่มีจุดศูนย์กลางร่วมกับแทรคอื่น ๆ เป็นจำนวนแสน ๆ วง แแทรคบันทึกข้อมูลเหล่านี้ยังถูกแบ่งย่อยเป็นท่อนข้อมูลหลาย ๆ ท่อน ตามความยาวของแทรคซึ่งแปรผันโดยตรงกับรัศมีของมัน ซึ่งเรียกว่า เซกเตอร์ (Sector) โดยทั่วไปนิยมใช้ขนาดของเซกเตอร์ที่บรรจุข้อมูลได้ 512 ไบท์ (byte) ดังแสดงในรูปที่ 2.3



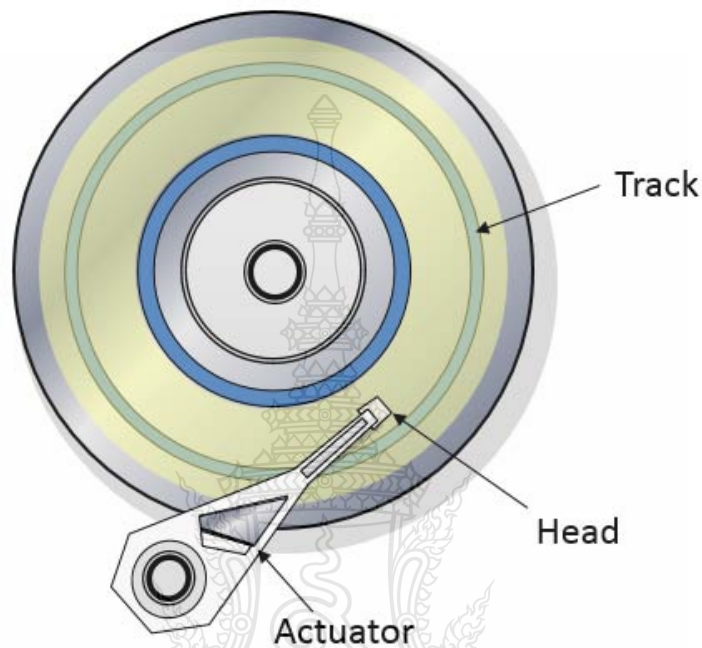
รูปที่ 2.3 การจัดแบ่งพื้นที่บนแผ่นจานแม่เหล็กสำหรับบันทึกข้อมูล [3]

โดยทั่วไปแทรควงนอกจะมีพื้นที่มากกว่าแทรควงใน ในขณะที่เซกเตอร์ในแต่ละแทรคจะมีจำนวนเท่ากัน ทำให้สิ้นเปลืองพื้นที่ในการจัดเก็บข้อมูล จึงมีการจัดรูปแบบใหม่ซึ่งเรียกเทคนิคนี้ว่าการจัดเก็บข้อมูลแบบแบ่งโซน ทำให้ฮาร์ดดิสก์สามารถเก็บข้อมูลได้เพิ่มขึ้น ดังแสดงในรูปที่ 2.4



รูปที่ 2.4 การจัดแบ่งพื้นที่บนแผ่นจานแม่เหล็กสำหรับบันทึกข้อมูลแบบแบ่งโซน [3]

2.1.3 แขนหัวอ่าน (Actuator Arm) ทำงานร่วมกับมอเตอร์ในการเลื่อนแขนหัวอ่านไปยังตำแหน่งที่เหมาะสมในการอ่าน/เขียนข้อมูลโดยมีคอนโทรลเลอร์ (Controller) ทำหน้าที่ในการแปลงคำสั่งที่มาจากคอมพิวเตอร์ จากนั้นก็เลื่อนหัวอ่านไปยังตำแหน่งที่ต้องการเพื่ออ่านหรือเขียนข้อมูลดังแสดงในรูปที่ 2.5



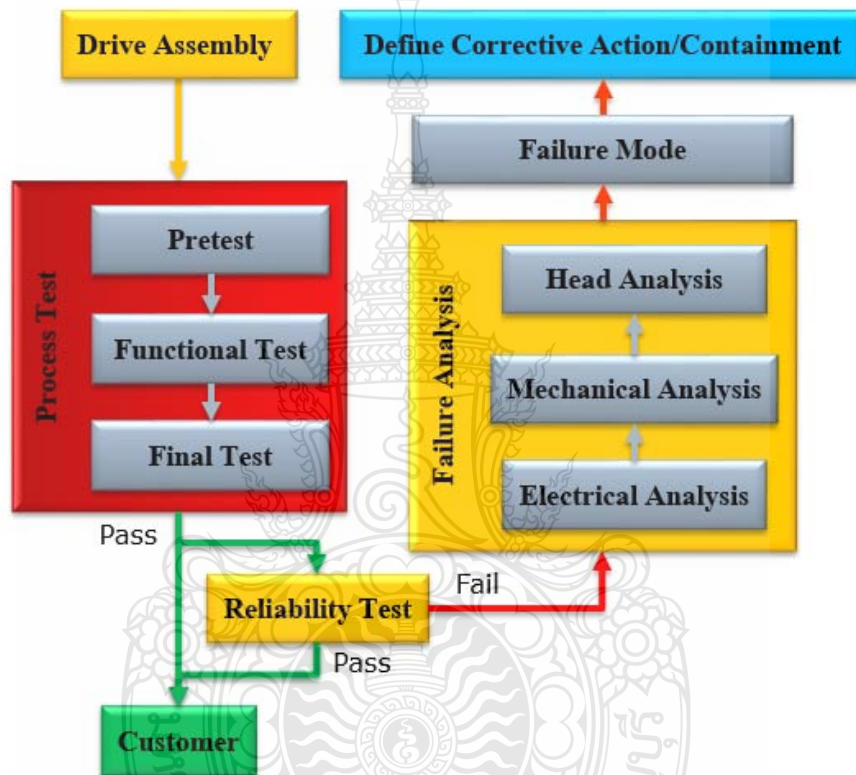
รูปที่ 2.5 การทำงานของแขนหัวอ่าน [3]

2.1.4 มอเตอร์หมุนจานแม่เหล็ก (Spindle Motor) ทำหน้าที่ควบคุมการหมุนของแผ่นจานแม่เหล็ก โดยทั่วไปฮาร์ดดิสก์จะมีความเร็วรอบอยู่ที่ประมาณ 5,400 หรือ 7,200 รอบต่อนาที และปัจจุบันฮาร์ดดิสก์ที่มีประสิทธิภาพสูงจะมีความเร็วรอบอยู่ที่ 10,000 รอบต่อนาที การพัฒนาให้ฮาร์ดดิสก์หมุนเร็วจะได้ประสิทธิภาพสูงขึ้น ซึ่งจะมีผลให้เซกเตอร์ที่บรรจุข้อมูลจะถูกหมุนมาพบกับหัวอ่านเร็วขึ้น ทำให้การค้นพบและอ่านข้อมูลได้เร็วขึ้น ประสิทธิภาพดีขึ้น เพราะเป็นลดเวลาในการเข้าถึงข้อมูล

2.1.5 กล่องใส่ฮาร์ดดิสก์ (Case) ใช้สำหรับบรรจุแผ่นจานแม่เหล็ก และกลไกในการหมุน การอ่าน ของฮาร์ดดิสก์ ผู้ผลิตแต่ละรายจะผลิตเคสไม่เหมือนกัน ปกติเคสจะถูกปิดผนึกอย่างแน่นหนา เพื่อป้องกันฝุ่นเข้าไปภายใน

## 2.2 มาตรฐานการทดสอบในอุตสาหกรรมฮาร์ดดิสก์

ในขั้นตอนการผลิตฮาร์ดดิสก์ประกอบด้วยกระบวนการต่าง ๆ ดังแสดงในรูปที่ 2.6 คือ Drive Assembly จะเป็นขั้นตอนการประกอบชิ้นส่วนของฮาร์ดดิสก์ ส่วน Pretest, Functional Test และ Final Test จะรวมอยู่ในกระบวนการทดสอบคุณภาพของกระบวนการผลิต และจะมีการสุ่มมาทดสอบความน่าเชื่อถือ อีกจำนวนหนึ่งก่อนส่งมอบให้ลูกค้า ซึ่งหากไม่ผ่านการทดสอบจะถูกส่งไปยัง Failure Analysis เพื่อวิเคราะห์ห้อย่างละเอียดเพื่อหาสาเหตุของอาการเสียและแก้ไขได้อย่างถูกต้อง



รูปที่ 2.6 กระบวนการทดสอบคุณภาพฮาร์ดดิสก์ที่ใช้งานในปัจจุบัน

มาตรฐานการทดสอบคุณภาพของฮาร์ดดิสก์สามารถแบ่งออกเป็น 2 ขั้นตอนที่สำคัญ คือ ทดสอบตามขั้นตอน และทดสอบความน่าเชื่อถือ โดยขั้นตอนทดสอบตามขั้นตอน [5] จะทำการทดสอบฮาร์ดดิสก์ในลักษณะต่าง ๆ และปรับแต่งค่าพารามิเตอร์เพื่อให้สามารถใช้งานได้เหมาะสมเป็นไปตามมาตรฐานอุตสาหกรรมฮาร์ดดิสก์ โดยมาตรฐานที่สำคัญได้มีขั้นตอนการทดสอบดังนี้

2.2.1 Product Setup เป็นการตั้งค่าเริ่มต้นให้ตัวฮาร์ดดิสก์ เช่น กำหนดจำนวน Head, Drive SN, PCBA SN, Family ลงใน Flash Memory

2.2.2 PCBA Screen เพื่อตรวจสอบ DRAM บน PCBA โดยการเขียนรูปแบบ (pattern) ลง DRAM และอ่านกลับพร้อมทั้งตรวจสอบข้อมูลที่อ่านออกมากับข้อมูลที่เขียนลงไป โดยการเขียนรูปแบบจะถูกเปลี่ยนค่าของแรงดันไฟฟ้าไปพร้อม ๆ กัน ค่าของแรงดันไฟฟ้าจะมีสามค่าคือ ต่ำ ปกติ และสูง จากนั้นเมื่อเสร็จจะทำการปรับค่าของแรงดันไฟฟ้ากลับคืน

2.2.3 Head Calibration ปรับจูนค่ากระแสไฟฟ้าของหัวอ่าน วัดค่าความต้านทาน MR (Magneto-Resistive) และ Heater

2.2.4 MDW Calibration ปรับจูนค่าชดเชยต่อ RRO เนื่องจากความเอียงของแผ่นจานแม่เหล็ก (Disc Slip) ได้

2.2.5 PES Screen เพื่อหาประสิทธิภาพในการควบคุมตำแหน่งหัวได้ดีเพียงใดบนแทรค โดยการวัด RRO (Repeatable Run Out) and NRRO (Non Repeatable Run Out) ของฮาร์ดดิสก์จากสัญญาณ PES ทั้งแผ่นจานแม่เหล็ก

2.2.6 Adaptive Fly Height ปรับจูนระบบการบินเบื้องต้นโดยใช้วิธีการวัดการสัมผัส โดยใช้ค่ามาตรฐานของ Write Power ในการหาค่าของระยะห่างระหว่างหัวอ่านและแผ่นจานแม่เหล็ก (Clearance) ซึ่งในการวัดหาค่าระยะห่างระหว่างหัวอ่านและแผ่นจานแม่เหล็ก นี้จะมีการวัด 2 ส่วนคือ Write/Heat และ Heat อย่างเดียว

2.2.7 MR Resistance Calibration หาค่าความต้านทานของหัวอ่านหลังจากการทำกรวัดการสัมผัส เพราะหัวอ่านอาจเกิดการเสื่อมเนื่องจากการขีดสีระหว่างหัวอ่านกับแผ่นจานแม่เหล็ก

2.2.8 Automatic Gain Control เพื่อหาระยะห่างระหว่างหัวอ่านกับหัวเขียน (Read to Write) เนื่องจากระยะห่างระหว่างหัวอ่านกับหัวเขียน

2.2.9 Read/Write Gap Calibration ปรับจูนระยะห่างด้านเวลาของหัวอ่านกับหัวเขียนเบื้องต้น เนื่องจากตำแหน่งของหัวอ่านกับหัวเขียนมีระยะทางระหว่างกัน จึงมีการเหลื่อมของเวลาเกิดขึ้น ต้องมีการปรับจูนความเหลื่อมทางด้านเวลาเหล่านี้เพื่อป้องกันไม่ให้เขียนทับรูปแบบเซอร์โวหรือข้อมูลที่อยู่รอบข้าง

2.2.10 Heater Induced Reader Protrusion ปรับจูนค่าสัมประสิทธิ์ของระบบการบิน ในการคำนวณรูปแบบการบินของหัวอ่านในช่วงแรกจะใช้ค่าสัมประสิทธิ์มาตรฐานมาใช้ในการคำนวณ

2.2.11 Adaptive Fly Height Screen คัดแยกฮาร์ดดิสก์ที่มีระบบการบินที่ไม่ดีโดยอาศัยค่าความแตกต่างของการบินที่วัดได้ระหว่างโซนข้อมูล

2.2.12 Head Stability ตรวจวัดความไม่เสถียรของหัวอ่าน โดยเขียนรูปแบบข้อมูลเป็น 00 แล้วเขียนทับด้วยรูปแบบข้อมูลต่าง ๆ ได้แก่ 00, AA และ FF แล้วปรับค่าชดเชยของแทรค (Offset track) ช่วง -192 to 192 เพื่อตรวจจับค่าความผิดพลาด

2.2.13 Variable Bit Aspect Ratio ปรับจูนหาความจุของฮาร์ดดิสก์ให้ได้ตามเป้าหมายของแต่ละผลิตภัณฑ์ โดยทั่วไปแล้วหัวอ่านแต่ละตัวจะมีประสิทธิภาพแตกต่างกัน จึงจำเป็นต้องหาค่า TPI (Track Per Inch) และ BPI (Bits Per Inch) แล้วนำมาหาค่าความหนาแน่น (Areal Density) ให้ได้ความจุตามเป้าหมาย

Areal Density (BPI\*TPI): ปริมาณของข้อมูลที่สามารถเก็บไว้ใน 1 ตารางนิ้ว

Bit Density: จำนวนบิตที่พอดีใน 1 นิ้ว ของแทรค (BPI)

Track Density: จำนวนแทรคที่พอดีใน 1 นิ้ว (TPI)

2.2.14 Read Screen คัดแยกประสิทธิภาพการอ่านของฮาร์ดดิสก์โดยการสุ่มทดสอบหาค่า BER (Bit Error Rate) เพื่อวัดประสิทธิภาพของหัวอ่านว่าทำงานได้ดีเพียงใด

ส่วนในขั้นตอนทดสอบความน่าเชื่อถือ [6] เป็นการทดสอบฮาร์ดดิสก์ในลักษณะการจำลองให้เป็นไปตามสภาวะต่าง ๆ ตามที่ลูกค้านำไปใช้งาน โดยใช้ระยะเวลาในการทดสอบประมาณ 2-3 วัน และด้วยข้อจำกัดของเครื่องที่ใช้ทดสอบทำให้ในแต่ละผลิตภัณฑ์สามารถสุ่มมาทดสอบได้เพียง 10-15% เท่านั้น ซึ่งการทดสอบนี้เพื่อให้มั่นใจได้ว่าเมื่อลูกค้านำไปใช้งานแล้วจะไม่เกิดปัญหา โดยมีลำดับการทดสอบที่สำคัญดังตารางที่ 2.1

## ตารางที่ 2.1 มาตรฐานที่สำคัญในกระบวนการทดสอบความน่าเชื่อถือของฮาร์ดดิสก์

การทดสอบ	รายละเอียด
1. Drive Power On	ทดสอบการความพร้อมการใช้งาน เป็นการตรวจวัดเวลาตั้งแต่การจ่ายไฟจนกระทั่งฮาร์ดดิสก์พร้อมใช้งาน
2. Write Read Perform	ทดสอบการอ่าน/เขียนข้อมูล
3. System File Load	ทดสอบการโหลดไฟล์ข้อมูลจาก S.M.A.R.T มาทำการวิเคราะห์
4. Voltage High Low	ทดสอบการจ่ายกระแสไฟสูงสุดและต่ำสุดตามมาตรฐาน เพื่อตรวจสอบการทำงานของฮาร์ดดิสก์สามารถอ่าน/เขียนได้ตามปกติ
5. Temperature High Low	ทดสอบการปรับอุณหภูมิสูงสุดและต่ำสุดตามมาตรฐาน เพื่อตรวจสอบการทำงานของฮาร์ดดิสก์สามารถอ่าน/เขียนได้ตามปกติ
6. Write Pattern	เขียนรูปแบบสนามแม่เหล็กให้มีการจัดเรียงทิศทางในลักษณะต่าง ๆ
7. Read Pattern	ตรวจสอบสนามแม่เหล็กที่มีการจัดเรียงทิศทางต่าง ๆ ตามที่เขียนไว้
8. Random Write	ทดสอบการสุ่มเขียนข้อมูลลงบนจานแม่เหล็กในแต่ละโซนข้อมูล
9. RAM Miss-compare	ทดสอบการทำงานของ RAM โดยการเปรียบเทียบ Buffer data
10. Write Zero Pattern	เขียนรูปแบบสนามแม่เหล็กให้มีการจัดเรียงทิศทางที่มีสถานะเป็น 0
11. Read Zero Verify	ตรวจสอบสนามแม่เหล็กที่มีการจัดเรียงทิศทางที่มีสถานะเป็น 0
12. Long Drive Self-Test	ทดสอบและตรวจสอบการทำงานด้วยตัวเองแล้วเก็บไว้ในไฟล์ข้อมูล
13. SMART Reset	ล้างข้อมูล S.M.A.R.T เพื่อพร้อมสำหรับการเริ่มต้นใช้งาน
14. Drive Power Off	ปิดการจ่ายกระแสไฟเพื่อให้ฮาร์ดดิสก์หยุดทำงาน

## 2.3 การวิเคราะห์อาการเสียของฮาร์ดดิสก์ (Failure Mode Analysis)

ในการวิเคราะห์สาเหตุการเสียของฮาร์ดดิสก์ [7] ที่ไม่ผ่านการทดสอบความน่าเชื่อถือนั้น จะต้องวิเคราะห์โดยผู้เชี่ยวชาญ โดยมีขั้นตอนการทดสอบตามลำดับขั้น ดังนี้ ทดสอบด้วยซอฟต์แวร์ทดสอบ (1st Verification Station) ซึ่งมีตรวจสอบข้อมูลจำเพาะของฮาร์ดดิสก์ จากนั้นจะทำการถอดรหัสข้อมูลของล็อกไฟล์ต่าง ๆ และการค้นหาตำแหน่งที่ผิดพลาดบนแผ่นจานแม่เหล็ก ในกระบวนการนี้จะใช้เวลาเฉลี่ย 6 ชั่วโมง จากนั้นผู้เชี่ยวชาญจะนำล็อกไฟล์ที่ได้มาทำการวิเคราะห์สาเหตุการเสียโดยการวิเคราะห์ข้อมูลของสมาร์ทจากแอดทริบิวท์ ซึ่งกระบวนการนี้จะใช้เวลาเฉลี่ย 30 นาที จากนั้นเมื่อทราบกลุ่มของประเภทการเสียแล้ว จะทำการทดสอบทางไฟฟ้า (Electrical Analysis) โดยกระบวนการนี้ใช้เวลาเฉลี่ย 3 ชั่วโมง เมื่อได้ผลการทดสอบแล้วจะทำการส่งฮาร์ดดิสก์

และผลการทดสอบที่ได้ไปยังผู้เชี่ยวชาญทางด้านกลไก (Mechanical Analysis) และผู้เชี่ยวชาญทางด้านส่วนประกอบ (Component Analysis) เช่น หัวอ่าน, แผลงวงจร และแผ่นจานแม่เหล็ก เป็นต้น ซึ่งใช้เวลาเฉลี่ย 4 ชั่วโมงและ 24 ชั่วโมงตามลำดับ จากนั้นเมื่อได้ผ่านการทดสอบด้วยกระบวนการทั้งหมดแล้ว จะเข้าสู่กระบวนการค้นหาวิธีการป้องกันปัญหาที่เกิดขึ้นใช้เวลาเฉลี่ย 48 ชั่วโมง โดยมีการแยกอาการเสียของฮาร์ดดิสก์ออกเป็นประเภทต่าง ๆ ที่สำคัญได้ 14 ประเภท ดังนี้

2.3.1 NMD (New Media Defect) คือ สาเหตุเกิดจากหัวอ่านไปสัมผัสกับแผ่นจานแม่เหล็กหรือหัวอ่านได้นำสิ่งปนเปื้อน ฝุ่นละอองต่าง ๆ (Contamination) [8] ที่อยู่ภายในฮาร์ดดิสก์ไปสัมผัสกับแผ่นจานแม่เหล็ก ทำให้แผ่นจานแม่เหล็กเป็นรอย (Media Scratch) ซึ่งสาเหตุนี้ทำให้ข้อมูลในบริเวณที่มีรอยเสียหาย ไม่สามารถกู้คืนและบันทึกข้อมูลบริเวณดังกล่าวได้

2.3.2 FHM (Fly Height Modulation) คือ อาการเสียลักษณะนี้เกิดจากการแกว่งของหัวอ่านขณะทำการเขียนข้อมูลไปตามแทรค ซึ่งอาจเกิดจากมีคลื่นความถี่รบกวน (Resonance) ขณะทำการเขียน หรืออ่านตำแหน่งที่จะเขียนบนแผ่นจานแม่เหล็กได้ไม่บ่อยดี

2.3.3 Head Instability เกิดจากการทำงานที่ผิดปกติของหัวอ่าน โดยจะแสดงออกมาในรูปของสัญญาณรบกวนซึ่งทำให้การอ่านข้อมูลผิดพลาดมากขึ้น เช่น Asymmetry Head เกิดจากหัวอ่านมีความไม่สามาร ทำให้อ่านสัญญาณข้อมูลจากแผ่นจานแม่เหล็กได้ไม่สมบูรณ์ Baseline Noise เกิดปัญหาที่หัวอ่านมีรอยแตกหักทำให้ค่าความต้านทานเปลี่ยนไป เมื่อทำการอ่านข้อมูลจากแผ่นจานแม่เหล็กก็จะทำให้มีสัญญาณผิดเพี้ยนไป

2.3.4 Weak Write คือ เกิดจากการเขียนที่ผิดพลาดในระยะเวลาหนึ่ง จัดเป็นลักษณะความผิดพลาดที่สามารถแก้ไขได้ ซึ่งเกิดจากปัจจัยต่าง ๆ ได้แก่ ระยะห่างระหว่างหัวอ่านกับแผ่นจานแม่เหล็กมากกว่าปกติ หรือเกิดจากสิ่งปนเปื้อน หรือฝุ่นละอองได้บดบังขณะที่หัวเขียนกำลังเขียนข้อมูลลงบนแผ่นจานแม่เหล็ก หรืออาจเกิดจากการปล่อยกระแสเข้ามาให้กับขดขวดมีน้อยเกินไป ทำให้มีสนามแม่เหล็กลดน้อยเกินไป หรือเกิดจากระยะห่างระหว่างหัวเขียนกับแผ่นจานแม่เหล็กมีระยะที่ไม่เหมาะสม ทำให้เวลาอ่านข้อมูลบริเวณนั้นก็จะได้ข้อมูลที่ขาดหายไปบางส่วน

2.3.5 VDW (Vertically Displayed Writes) คือ มีลักษณะอาการคล้ายกับ Weak Write โดยข้อมูลจะหายไปประมาณ 1-2 เซกเตอร์ ซึ่งมีลักษณะการเขียนที่ไม่สม่ำเสมอเกิดจากการที่หัวอ่านไม่สามารถเขียนบริเวณตำแหน่งนั้น ได้เนื่องจากอาจมีฝุ่นผงบดบังหัวอ่านในส่วนของการเขียนหรือเกิดจากความไม่คงที่ของระยะห่างระหว่างหัวอ่านกับแผ่นจานแม่เหล็ก



2.3.6 Degraded Reader คือ หัวอ่านข้อมูลมีปัญหา เช่น ค่าความต้านทาน (MR Degraded) สัญญาณรบกวน (Noise) หรือ หัวอ่านคดงอ (Head Bent) และปัญหาที่เกิดจากฝุ่นละออง เป็นต้นซึ่งสาเหตุเหล่านี้ทำให้ไม่สามารถอ่านข้อมูลจากแผ่นจานแม่เหล็กได้

2.3.7 Degraded Writer คือ หัวเขียนข้อมูลมีปัญหา อาจเกิดจากการที่หัวอ่านในส่วนที่ใช้เขียนข้อมูลมีการชนกันกับพื้นผิวของแผ่นจานแม่เหล็กที่นูนขึ้นมาซึ่งเกิดจากความร้อน หรือ TA (Thermal Asperity) ที่อยู่บนจานแม่เหล็ก อาจทำให้หัวอ่านในส่วนของการเขียนมีรอย จึงทำให้การทำงานผิดปกติ หรืออาจทำให้ความต้านทานของหัวอ่านที่มีการเปลี่ยนแปลงไปจากเดิม รวมถึงปัจจัยจากสภาพแวดล้อมอื่น ๆ กล่าวคือเมื่อมีการเขียนข้อมูล จะทำให้ไม่สามารถเขียนข้อมูลลงบนแผ่นจานแม่เหล็กได้

2.3.8 Magnetic Particle Erasure คือ เกิดจากการลบข้อมูลบน สนามแม่เหล็กในขณะที่หัวอ่านบินผ่านบริเวณนั้นทำให้การจัดเรียงสนามแม่เหล็กมีความผิดปกติแต่สามารถกู้ข้อมูลคืนได้

2.3.9 OTF (Off-track Write) คือ เกิดจากความผิดพลาดในขณะที่บันทึกข้อมูลมีแรงกระทำจากภายนอก (Operation Shock) [9-10] หรือการกำทอนจากอุปกรณ์ภายใน ทำให้ข้อมูลที่เขียนไม่ถูกต้องและอาจมีผลทำให้แทรคที่อยู่ใกล้เคียง

2.3.10 ATI คือ แทรคข้างเคียงโดนสนามแม่เหล็กจากแทรคที่กำลังเขียนรบกวนในขณะที่เขียนข้อมูลลงบนแผ่นจานแม่เหล็ก

2.3.11 STE คือ อาการคล้าย ATI แต่จะเกิดกับแทรคข้างเคียงที่ถัดไป 2-3 แทรค

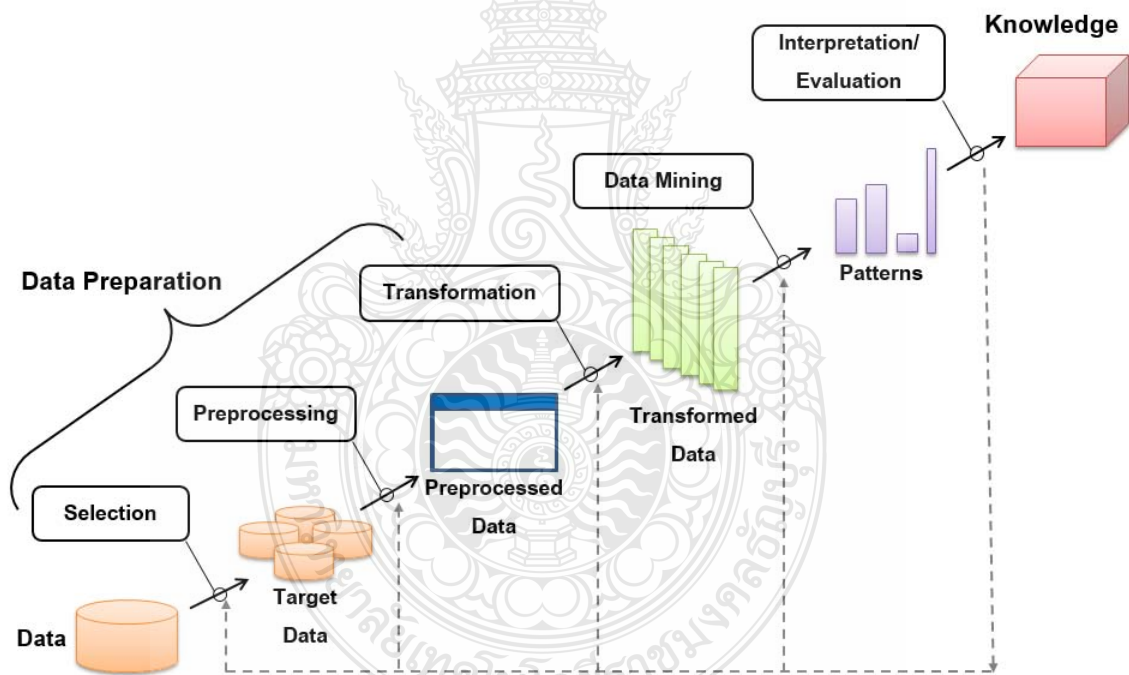
2.3.12 Aborted Write คือ เกิดจากปัจจัยภายนอกมากระทำในขณะที่มีการเขียนข้อมูล หรือการจ่ายกระแสที่ไม่สม่ำเสมอ ทำให้ถูกสั่งให้หยุดเขียนอย่างกะทันหัน

2.3.13 CND (Can Not Duplicate) คือ ฮาร์ดดิสก์อาจเคยมีปัญหามาก่อนแต่สามารถแก้ไขข้อผิดพลาดได้อัตโนมัติ และเมื่อมีการนำมาทดสอบหาข้อผิดพลาดจากจุดที่มีการบันทึกไว้ จะไม่สามารถพบข้อผิดพลาดเนื่องจากถูกระบบการแก้ไขไปแล้ว ก็จะไม่ทราบสาเหตุที่เคยเกิดขึ้น

2.3.14 NPF (No Problem Found) คือ ไม่มีการบันทึกจุดที่มีปัญหา และไม่พบสิ่งผิดปกติ โดยอาจจะทดสอบโดยเครื่องมือต่าง ๆ เพื่อแสดงว่าไม่พบความผิดปกติจากตำแหน่งใด ๆ บนแผ่นจานแม่เหล็ก ซึ่งรูปแบบนี้จะสามารถนำไดร์ฟกลับไปใช้ได้ตามปกติ

## 2.4 การทำเหมืองข้อมูล (Data Mining)

2.4.1 กระบวนการทำงานของเหมืองข้อมูล [11] คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น สามารถดำเนินการได้โดยอัตโนมัติ ทำให้มีชื่อเรียกอีกชื่อหนึ่งว่า KDD (Knowledge discovery in database) ความสามารถในการสังเคราะห์ความรู้ขึ้นได้เองทำให้ระบบเหมืองข้อมูลเหมาะสมที่จะถูกนำไปใช้ช่วยในระบบสนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพสูง จึงได้รับความนิยมในการนำไปใช้อย่างแพร่หลาย ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม ข้อมูลที่สังเคราะห์โดยระบบเหมืองข้อมูลแสดงผลออกมาได้หลายรูปแบบ เช่น แสดงเป็นภาพ แสดงเป็นกราฟ หรือเป็นสมการคณิตศาสตร์ ขึ้นอยู่กับลักษณะงานของแต่ละประเภท ซึ่งโดยทั่วไปกระบวนการของ KDD นั้นประกอบด้วยขั้นตอนต่าง ๆ ดังนี้



รูปที่ 2.7 กระบวนการทำงานของเหมืองข้อมูล [11]

จากรูปที่ 2.7 กระบวนการทำงานของเหมืองข้อมูลประกอบด้วยขั้นตอนการทำงานหลักอยู่ 4 ขั้นตอน สามารถอธิบายขั้นตอนต่าง ๆ ได้ ดังต่อไปนี้

2.4.1.1 Business Object ขั้นตอนแรกจะต้องมีการกำหนดขอบเขต เป้าหมายในการทำเหมืองข้อมูลก่อน ว่าใช้เพื่อการตัดสินใจเรื่องอะไร ซึ่งขั้นตอนนี้จะมีผลต่อทุก ๆ ขั้นตอนในการทำ

เหมืองข้อมูล โดยนักวิเคราะห์จะต้องมีการกำหนดปัญหาที่เกิดขึ้นหรือความต้องการทางธุรกิจให้ครอบคลุม ชัดเจน และกำหนดวัตถุประสงค์ด้วย เช่น ต้องการ เพิ่มยอดการตอบรับบริการขายทางจดหมาย ขึ้นอยู่กับการระบุเป้าหมายว่า จะเพิ่มอัตราการตอบรับหรือเพิ่มมูลค่าการตอบรับซึ่ง จำ เป็นที่จะต้องสร้าง Model ที่แตกต่างกัน วัตถุประสงค์ที่กำหนดขึ้นมาจะต้องมีการระบุวิธีการในการวัดผลลัพธ์ที่ได้จาก โครงการ รวมถึงต้นทุนที่สมเหตุสมผลด้วย

2.4.1.2 Data Preparation ในการวิเคราะห์ข้อมูลที่มีขนาดใหญ่จำเป็นต้องมีการลดข้อมูลที่ใช้ โดยการใช้เทคนิคต่าง ๆ เข้ามาช่วยเลือกและปรับแต่งข้อมูล [12] ที่มีความสำคัญต่อการวิเคราะห์ สามารถจำแนกได้เป็นประเภทดังต่อไปนี้

1) Data Selection เป็นการคัดเลือกข้อมูลที่เกี่ยวข้องกับเป้าหมาย หรือ วัตถุประสงค์ที่ได้ตั้งไว้ เพื่อให้ผลการวิเคราะห์ข้อมูลหรือผลของการทำกระบวนการเหมืองข้อมูลสามารถนำไปใช้ในการประกอบการตัดสินใจได้ตามความต้องการ โดยสามารถเลือกจากแหล่งข้อมูลต่าง ๆ ซึ่งแหล่งข้อมูลหลักของการทำเหมืองข้อมูล คือ คลังข้อมูล (Data Warehouse) หรือฐานข้อมูล (Database) ตัวแปรที่ถูกเลือกมาแต่ละตัวนั้นจะต้องถูกทำความเข้าใจว่าตัวแปรแต่ละตัวหมายความว่าอะไร ประกอบด้วยอะไร ไม่เพียงแต่คำ จำ กัดความทางธุรกิจเท่านั้น แต่จะต้องมีคำ อธิบายอย่างชัดเจนเกี่ยวกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดของข้อมูล รูปแบบของข้อมูล และลักษณะอื่น ๆ จะมีตัวแปร 2 ชนิด คือ

(1) ตัวแปรแบบ Categorical มี 2 ลักษณะ คือ 1) Nominal Variable กล่าวถึงชนิดนี้ของ Object ที่จะอ้างถึงแต่ไม่มีลำดับในค่าที่เป็นไปได้ (Possible Value) ตัวอย่างเช่น สถานะภาพแต่งงาน (โสด แต่งงาน หย่า ไม่ทราบ) เพศ (ชาย หญิง) ระดับการศึกษา (ปริญญาโท ปริญญาตรี ม.ปลาย ปวช) และ 2) Ordinal Variable มีลำดับสำหรับค่าที่เป็นไปได้ ตัวอย่างเช่น ลำดับของ ลูกค้า (ดี ปานกลาง ไม่ดี)

(2) ตัวแปรแบบ Quantitative ซึ่งมีการวัดความแตกต่างระหว่างค่าที่เป็นไปได้ มี 2 ลักษณะ คือ 1) Continuous (ค่าที่ต่อเนื่อง) เช่นรายได้ เฉลี่ยจำนวนครั้งที่ซื้อ รายได้ และ 2) Discrete (ค่าเป็นจำนวนเต็ม) เช่นจำนวนพนักงาน เวลาปี (เดือน ฤดู ไตรมาส)

2) Data Preprocessing เป็นขั้นตอนการปรับข้อมูลที่ได้คัดเลือกมานั้น ให้มีคุณภาพมากยิ่งขึ้น เนื่องจากข้อมูลส่วนใหญ่ที่มีในองค์กร ไม่ได้ถูกเตรียมมาเพื่องาน Data Mining โดยเฉพาะ และการคัดเลือกข้อมูลจากหลากหลายแหล่งข้อมูล อาจจะทำให้มีรูปแบบ (Format) ในการเก็บข้อมูลไม่เหมือนกัน ส่งผลให้เกิดการผิดพลาดในการทำเหมืองข้อมูลได้ ซึ่งการปรับข้อมูลในขั้นตอน Data Preprocessing ประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

(1) Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป หรือเติมข้อมูลที่ขาดหายไปบางส่วน รวมทั้งเลือกข้อมูลที่จำเป็นต่อการทำเหมืองข้อมูล เช่น การตัดทิ้งรายการที่มีข้อมูลสูญหาย การเติมค่าที่ขาดหายด้วยมือ การเติมค่าคุณลักษณะของข้อมูลที่ขาดหายทุกค่าด้วยค่าคงที่ค่าหนึ่ง การใช้ค่าเฉลี่ยของคุณลักษณะเติมค่าข้อมูลที่ขาดหาย การใช้ค่าเฉลี่ยคุณลักษณะของตัวอย่างที่จัดอยู่ในประเภทเดียวกัน การใช้ค่าที่เป็นไปได้มากที่สุด เติมแทนค่าข้อมูลที่ขาดหาย เช่น ค่าที่ได้จากสมการความถดถอย (Regression) ค่าที่ได้จากการอนุมาน โดยใช้สูตรของเบย์ (Bayesian formula) หรือต้นไม้ตัดสินใจ (Decision tree)

(2) Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน การผสานข้อมูลจากแหล่งต่าง ๆ เพื่อช่วยลดหรือหลีกเลี่ยงความซ้ำซ้อนของข้อมูล (Data Redundancies) ซึ่งจะนำไปสู่ปัญหาความไม่สอดคล้องกันของข้อมูล (Data inconsistencies) เช่น ข้อมูลที่มีรูปแบบการจัดเก็บต่างกัน หรืออาจมีหน่วยวัดที่ต่างกัน สัญลักษณ์ที่ใช้ไม่เหมือนกัน

(3) Data Reduction เป็นขั้นตอนการลดรูปของข้อมูล เพื่อแทนข้อมูลด้วยรูปแบบกะทัดรัดกินเนื้อที่น้อยกว่า แต่สามารถนำไปใช้วิเคราะห์แล้วเกิดผลลัพธ์ที่เท่ากันหรือเทียบเท่ากับการวิเคราะห์จากข้อมูลทั้งหมด วิธีในการลดรูปข้อมูลได้แก่ 1) Data Aggregation เป็นการลดรูปข้อมูลที่มักพบใน Data Cube ซึ่งเป็นแบบจำลองข้อมูลหลายมิติ (Multidimensional data model) 2) Numerosity Reduction การลดรูปข้อมูลโดยใช้เทคนิคนี้มี 2 ประเภทคือ Parametric Methods เป็นการสร้างแบบจำลองข้อมูลด้วยวิธีความถดถอยซึ่งใช้เทคนิคทางสถิติในการประมาณค่าตัวแปรตาม Non-Parametric Methods วิธีนี้ไม่ได้สร้างแบบจำลองข้อมูล แต่จะใช้เครื่องมือ อาทิเช่น แผนภาพฮิสโตแกรม (Histogram) หรือวิธีการจัดกลุ่ม (Clustering) แสดงการกระจายของข้อมูล และเก็บค่าตัวแทนกลุ่มแทนค่าข้อมูลจริง หรืออาจใช้วิธีทางสถิติ เช่น การสุ่มตัวอย่าง (Sampling) 3) Discretization and Concept Hierarchy Generation การทำให้ข้อมูลเป็นแบบไม่ต่อเนื่อง เป็นการลดรูปข้อมูลเลขจำนวนจริงโดยแบ่งค่าข้อมูลดิบออกเป็นช่วง แล้วแทนค่าแต่ละช่วงด้วยค่าฉลากช่วง (Interval Label) เนื่องจากอัลกอริทึมการทำเหมืองข้อมูลบางอย่าง รับเฉพาะข้อมูลแบบ Categorical และสามารถลดรูปข้อมูล Categorical หรือข้อมูลที่ไม่ต่อเนื่อง ด้วยการสร้างลำดับชั้นแนวคิด โดยการนิยามลำดับการเทียบ (mapping) กลุ่มแนวคิดในระดับล่าง ไปสู่ แนวคิดในระดับที่สูงขึ้น ซึ่งแนวคิดเป็นค่าที่ใช้บรรยายลักษณะกลุ่มของข้อมูล ข้อมูลดิบ

3) Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน ซึ่งมีวิธีการแปลงข้อมูลรูปแบบต่าง ๆ ดังนี้

(1) Normalization คือ การแปลงข้อมูลให้อยู่ในช่วงค่าที่กำหนด จัดได้ว่าเป็นการลดระยะห่างของข้อมูล เนื่องจากข้อมูลชนิดเดียวกัน แต่มีระยะห่างมากเกินไป ซึ่งทำให้มีผลต่อการวิเคราะห์ข้อมูลในบางโมเดลของการทำเหมืองข้อมูล เช่น ข้อมูลที่มีค่าช่วง 0-1,000,000 แปลงให้เป็นช่วง 0-20 เป็นต้น

(2) Discretization คือ กระบวนการแปลงข้อมูลที่เป็นจำนวนจริงต่อเนื่องให้เป็นช่วงข้อมูล จัดได้ว่าเป็นกำหนดกลุ่มให้ข้อมูล ลดการกระจายของข้อมูล เช่น กำหนดกลุ่มอายุของลูกค้า เช่น ช่วงอายุ 20-25 ปี ช่วงอายุ 26-30 ปี ช่วงอายุ 31-35 ปี เป็นต้น

(3) 1 of N coding คือ กระบวนการแปลงข้อมูลจากตัวอักษรให้เป็นตัวเลขที่ไม่ได้บอกถึงลำดับความมากนัก เป็นการกำหนดรหัสให้กับข้อมูล ตัวอย่างเช่น ชนิดของรถ Ford Lincoln Nissan ให้เป็น 100 010 001 ปกติแบบนี้มักจะเป็น Input ของ Neural Network

(4) Generalization คือ กระบวนการแปลงข้อมูลที่มีรายละเอียดมากเกินไป ให้มีความกระชับมากยิ่งขึ้น เช่น ที่อยู่ของลูกค้า ทำการแปลงข้อมูลให้กระชับขึ้นเป็นอยู่ในกรุงเทพฯ หรืออยู่ต่างจังหวัด เป็นต้น

2.4.1.3 Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่ ผลลัพธ์ที่ได้จากขั้นตอนนี้จะเป็นรูปแบบของความสัมพันธ์ของข้อมูลที่จะนำมาใช้ในการพยากรณ์ (prediction) หรือวิเคราะห์ต่อไป ซึ่งแต่ละเทคนิคจะมีอัลกอริทึมต่าง ๆ ให้เลือกใช้ ดังต่อไปนี้

1) Database Segmentation (Clustering) เป็นกระบวนการแบ่ง Database ออกเป็นกลุ่มเพื่อให้ง่ายต่อการวิเคราะห์ เช่นการแบ่งลูกค้าออกตามอายุ เพศ รายได้ เป็นต้น

2) Predictive Modeling เป็นการสร้างแบบจำลองพยากรณ์ แบ่งเป็น 2 ลักษณะ ดังนี้

(1) Classification เป็นการจัดกลุ่มให้กับแต่ละข้อมูลในฐานข้อมูล โดยมีการระบุค่า หรือลักษณะที่เป็นไปได้ของข้อมูลภายในแต่ละกลุ่ม เช่น การจัดกลุ่มของผู้ป่วยตามผลของการใช้ยารักษา เพื่อระบุรูปแบบการรักษาให้กับผู้ป่วยใหม่ ที่เข้ารับการรักษา เป็นต้น

(2) Value Prediction เป็นการพยากรณ์ค่าที่เป็นไปได้หรือการกระจายของค่าที่เป็นไปได้ของตัวแปรใด ๆ ในกลุ่มข้อมูล การทำนายค่าที่เป็นตัวเลข เช่น การทำนายภาษีที่จะเก็บได้ในปี เป็นต้น

3) Link Analysis (Associations) เป็นการหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูลเพื่อใช้ลักษณะของข้อมูลหนึ่ง ๆ ในการบอกลักษณะที่จะเกิดขึ้นกับข้อมูลอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูลในกลุ่มเดียวกัน เช่น การระบุว่าในกลุ่มของลูกค้าที่ซื้อนม

นั้น จะมีลูกค้า 64% ที่ซื้อขนมปังด้วย หรืออาจจะเป็นการหาความสัมพันธ์ของตัวแปรระหว่างกลุ่มข้อมูลก็ได้ เช่น ในทุก ๆ ครั้งที่ดัชนีของตลาดหุ้นหนึ่งลดลง 5% ดัชนีของตลาดหุ้นอื่นจะเพิ่มขึ้น 13% ภายในช่วง 2-6 เดือนหลังจากนั้น เป็นต้น ซึ่งลักษณะของการหาความสัมพันธ์นั้นอาจแบ่งได้เป็น 3 กลุ่ม คือ การหาความสัมพันธ์ระหว่างข้อมูล (Association discovery) การหาความสัมพันธ์ในลักษณะที่เป็นลำดับของข้อมูล (Sequential Pattern discovery) และ การหาความสัมพันธ์ของข้อมูลกับช่วงเวลาใด ๆ (Similar Time Sequence discovery)

4) Deviation Detection เป็นเทคนิคที่ใช้ในการแสดงลักษณะของข้อมูลที่ผิดปกติ หรือผิดไปจากที่คาดไว้ โดยมีการแสดงผลอยู่ในลักษณะที่สามารถทำความเข้าใจและแปลความหมายได้ง่าย เช่น การใช้กราฟ เป็นต้น

2.4.14 Interpretation/Evaluation หรือ Knowledge Presentation เป็นขั้นตอนสุดท้ายสำหรับนักวิเคราะห์ข้อมูล ที่จะต้องเก็บผลลัพธ์ของกระบวนการเหมืองข้อมูลมาทำการสรุปความหมาย หรือตีความของผลลัพธ์ที่ได้ ออกมาเป็นข้อมูลความรู้ใหม่ (Knowledge) ที่สามารถนำไปเป็นสารสนเทศที่ช่วยในการตัดสินใจของผู้บริหารต่อไปได้

ในกระบวนการ KDD หรือการทำเหมืองข้อมูลนั้น จะใช้เวลาในการเตรียมข้อมูลสำหรับการทำเหมืองข้อมูลมากถึง 60% เนื่องจากว่าผลการวิเคราะห์เพื่อให้ได้ความรู้ใหม่ หรือสารสนเทศที่ซ่อนเร้นอยู่ในข้อมูลเป็นจำนวนมากนั้น จะมีความถูกต้องมากน้อยแค่ไหน ขึ้นอยู่กับว่าข้อมูลที่นำไปวิเคราะห์นั้นมีความสอดคล้อง และมีความผิดพลาดเท่าใด ดังนั้นจึงต้องให้ความสำคัญกับการเตรียมข้อมูล ส่วนการทำเหมืองข้อมูลด้วยอัลกอริทึมต่าง ๆ นั้นใช้เวลาเพียง 10% เท่านั้น

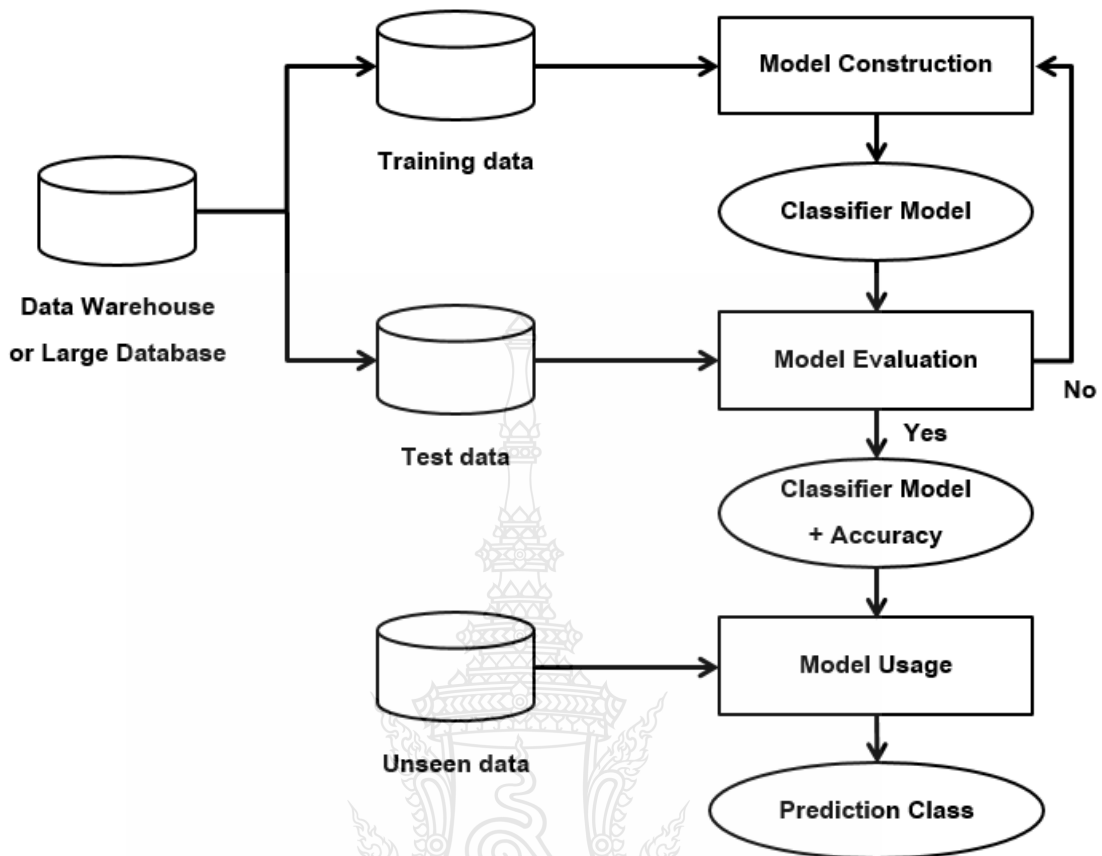
#### 2.4.2 เทคนิคการทำเหมืองข้อมูล สามารถแบ่งได้เป็น 2 กลุ่ม ดังนี้

2.4.2.1 การสร้างตัวแบบในการทำนาย (Predictive Modeling) หรือเรียกว่า (Supervised Learning) คือ การนำข้อมูลในอดีตมาสร้างต้นแบบเพื่อการทำนายอนาคต โดยมีการใช้ข้อมูลในการจัดทำตัวแบบ (Training Data) ซึ่งทุกข้อมูลจะมีคุณสมบัติ ซึ่งเป็นค่าที่ใช้ในการทำนายผลของข้อมูล อัลกอริทึมประเภทนี้จะมุ่งเน้นในการแบ่งแยกข้อมูลออกเป็นกลุ่มตามค่าคุณสมบัติของข้อมูล ซึ่งถ้าค่าคุณสมบัติของข้อมูลมีค่าไม่ต่อเนื่อง จะเรียกกระบวนการที่ใช้แบ่งแยกว่า การจำแนกประเภท (Classification) ถ้าค่าคุณสมบัติของข้อมูลมีค่าต่อเนื่อง จะเรียกกระบวนการที่ใช้ว่า การถดถอย (Regression) หรือการพยากรณ์ (Forecasting)

2.4.2.2 การสร้างตัวแบบในการบรรยาย (Descriptive Modeling) หรือเรียกว่า (Unsupervised Learning) คือ การนำข้อมูลที่มีอยู่มาดูและศึกษาเพื่อหากฎความสัมพันธ์ต่าง ๆ

(Association Rules) หรือหาการจัดกลุ่มข้อมูล (Clustering Analysis) ซึ่งไม่ได้มีจุดมุ่งหมายเพื่อการทำนาย

ในงานวิจัยนี้จะเน้นการศึกษาเฉพาะงาน การจำแนกประเภท (Classification) เนื่องจากใช้ประโยชน์ได้โดยตรงกับงานด้านการทำนายอาการของฮาร์ดดิสก์ การทำเหมืองข้อมูลประเภท Classification เป็นการจัดแบ่งประเภทของข้อมูล จะแบ่งการทำงานออกเป็น 2 ช่วง คือ ช่วงสร้างโมเดล และช่วงทดสอบโมเดล โดยหาชุดต้นแบบหรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์เพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูลที่ไม่มีการระบุประเภทหรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น และจะนำข้อมูลส่วนที่เหลือจาก Training Data เป็นข้อมูลที่ใช้ทดสอบ (Testing Data) ซึ่งเป็นกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลเพื่อทดสอบความถูกต้อง แล้วปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ หลังจากนั้นเมื่อมีข้อมูลใหม่เข้ามา ก็จะนำข้อมูลผ่านโมเดล โดยโมเดลจะสามารถทำนายกลุ่มของข้อมูลนี้ได้ ตัวอย่างของการจัดหมวดหมู่ เช่น การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำระดับกลาง และระดับสูงของความเสียหายที่จะได้รับ เป็นต้น ซึ่งกระบวนการสร้างตัวโมเดลจำแนกประเภทข้อมูล แบ่งออกเป็น 3 ขั้นตอน ซึ่งภาพรวมของกระบวนการสร้างโมเดลจำแนกประเภทข้อมูล [13] ดังแสดงในรูปที่ 2.14



รูปที่ 2.8 ขั้นตอนการสร้างโมเดล การทดสอบโมเดล และการใช้โมเดลเป็นตัวจำแนกข้อมูล [13]

จากรูปที่ 2.8 อธิบายกระบวนการคิดแยกแต่ละขั้นตอนได้ดังนี้

1) Model Construction (Learning) เป็นขั้นตอนการสร้างโมเดลจำแนกประเภท โดยอาศัยการเรียนรู้จากข้อมูลที่ได้กำหนด class ไว้เรียบร้อยแล้วหรือเรียกว่าข้อมูลเรียนรู้ (Training data) ซึ่งโมเดลจำแนกประเภทที่ได้จะแสดงด้วยวิธีการพื้นฐานทางเหมืองข้อมูลยกตัวอย่าง เช่น ต้นไม้ตัดสินใจโมเดลจำแนกประเภทที่ได้จะมีลักษณะคล้ายต้นไม้จริงกลับหัวที่มีโหนดรากอยู่ด้านบนสุดและโหนดใบอยู่ด้านล่างสุดของต้นไม้ แต่ละโหนดบนต้นไม้จะมีแอตทริบิวต์เป็นตัวเลือกทดสอบ ซึ่งจะมีกิ่งซึ่งเป็นค่าที่เป็นไปได้ของแอตทริบิวต์ที่ถูกเลือกทดสอบไว้ และมีโหนดใบแสดงกลุ่มที่กำหนดไว้

2) Model Evaluation (Accuracy) เป็นขั้นตอนตรวจสอบความถูกต้อง โดยอาศัยข้อมูลที่ใช้สำหรับทดสอบเรียกว่าข้อมูลทดสอบ (Testing data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบจะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลจำแนกประเภท เพื่อทดสอบว่าโมเดล



จำแนกประเภทนี้สามารถจัดกลุ่มประเภทข้อมูลได้อย่างถูกต้องมากน้อยเพียงใด และมีการปรับปรุงโมเดลจำแนกประเภทจนกว่าจะได้ค่าความถูกต้องในระดับที่ยอมรับได้

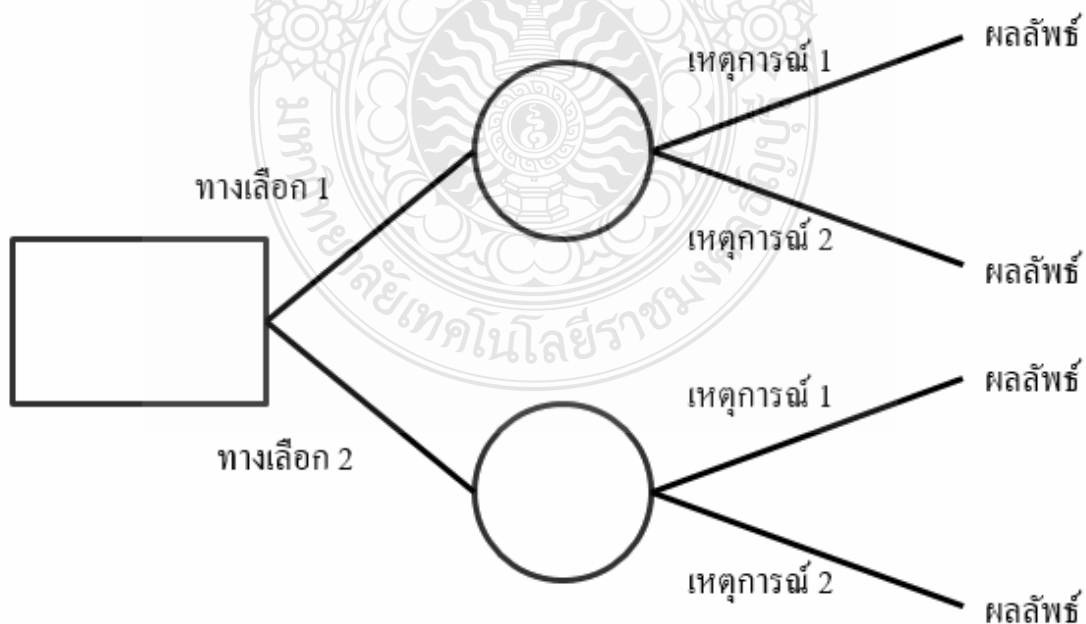
4) Model Usage (Classification) เป็นขั้นตอนการนำโมเดลจำแนกประเภทที่สร้างขึ้นมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน (unseen data) เพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น

### 2.4.3 อัลกอริทึมที่ใช้ในการทำเหมืองข้อมูล

การแก้ปัญหาของงานชนิดต่าง ๆ โดยใช้วิธีการเหมืองข้อมูลในแต่ละงานก็จะมีเทคนิคของการทำเหมืองข้อมูลที่จะนำมาใช้ได้อย่างเหมาะสม ซึ่งส่วนใหญ่จะเป็น AI (Artificial Intelligence) เทคนิคที่ใช้กันทั่วไป ได้แก่ Classification Rules, Decision Trees หรือ Neural Networks ซึ่งในงานวิจัยนี้จะเน้นการศึกษาเฉพาะเทคนิค Decision Trees ชนิดอัลกอริทึม C5.0

#### 2.4.3.1 ต้นไม้ตัดสินใจ (Decision Trees)

โครงสร้างแบบต้นไม้ตัดสินใจ เป็นแบบจำลองที่มีลักษณะคล้ายกับต้นไม้ [14] จะมีการสร้างกฎต่าง ๆ ขึ้นเพื่อใช้ในการตัดสินใจ ต้นไม้ตัดสินใจเป็นวิธีที่ได้รับความนิยมเนื่องจากความไม่ซับซ้อนของอัลกอริทึม ทำให้เครื่องมือที่ใช้ในการทำที่วางขายกันอยู่ในท้องตลาดต่างก็ใช้วิธีนี้ ข้อดีของวิธีนี้คือ สามารถตีความและเข้าใจลักษณะของรูปแบบข้อมูลได้ง่าย เพราะมีการจำแนกประเภทข้อมูลออกเป็นกลุ่มต่าง ๆ โดยใช้แอตทริบิวท์ของข้อมูลในการจำแนกประเภท แต่ก็ยังคงมีปัญหาในเรื่องของการให้ค่าน้ำหนักในแต่ละโหนดซึ่งถ้าให้น้ำหนักผิดไป อาจจะทำให้การตีความผิดไปได้ ส่วนประกอบของต้นไม้ตัดสินใจ แสดงได้ดังในรูปที่ 2.9



รูปที่ 2.9 ส่วนประกอบของต้นไม้ตัดสินใจ [14]

(1) โหนดราก คือ เป็นจุดเริ่มต้นของต้นไม้ตัดสินใจ ซึ่งเมื่อข้อมูลใด ๆ ตก  
ลงมาที่โหนด จะใช้คุณลักษณะนี้เป็นตัวตัดสินใจว่าข้อมูลจะไปในทิศทางใด

(2) กิ่ง คือ ค่าของคุณลักษณะในโหนดรากที่แตกกิ่งนี้ออกมา ซึ่งโหนด  
รากจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าของคุณลักษณะในโหนดรากนั้น

(3) โหนดใบ คือ กลุ่มต่าง ๆ ซึ่งเป็นผลลัพธ์ในการจำแนกประเภทข้อมูล  
กระบวนการคัดเลือกโหนดโดยใช้ฟังก์ชันเกน (Gain Function) เป็นการ  
จำแนกประเภทของข้อมูลให้อยู่ในประเภทเดียวกันในแต่ละโหนดมากที่สุด มีการตัดสินใจเลือก  
คุณสมบัติที่จะใช้เพื่อเป็นโหนดเริ่มต้น และ โหนดอื่น ๆ โดยใช้ความรู้จากทฤษฎีสารสนเทศ (Information  
Theory) [15] ซึ่งขึ้นอยู่กับความน่าจะเป็นของข้อมูลในการเลือกคุณสมบัติที่จะมาเป็น โหนดรากนั้น  
จะมีการคำนวณจากค่าเอนโทรปีทั้งหมดของชุดข้อมูลนั้นลบด้วยค่าเอนโทรปีหลังจากการเลือก  
คุณสมบัติใดคุณสมบัติหนึ่งเป็นราก ค่าเอนโทรปีหลังจากแบ่งตามคุณสมบัติที่เลือกแล้วสามารถ  
คำนวณได้จาก ค่าผลรวมของผลคูณระหว่างค่าเอนโทรปีของแต่ละโหนดกับอัตราส่วนของตัวอย่าง  
ในแต่ละกิ่งต่อตัวอย่างทั้งหมดที่โหนดนั้น ๆ จากนั้นเมื่อนำข้อมูลผ่านกระบวนการต่าง ๆ จนได้  
แบบจำลองที่สามารถจำแนกประเภทของข้อมูลได้แล้ว จะถูกแปลงเป็นกฎในการสร้างต้นไม้ตัดสินใจ  
ด้วยวิธีการกำหนดเงื่อนไขตามเส้นทางที่เริ่มจากโหนดรากไปจนถึงโหนดสุดท้ายของแต่ละกิ่ง โดยจะ  
เลือกแอตทริบิวท์ที่ให้ค่า Gain สูงสุด ซึ่ง Gain เป็นค่าที่บอกระดับความสามารถของการจำแนก  
ประเภทของแอตทริบิวท์ สามารถคำนวณได้ตามสมการด้านล่าง [16] ดังนี้

$$\text{Gain}(x) = \text{info}(T) - \text{info}_x(T) \quad (2.1)$$

เมื่อ  $T$  คือ เซตของ Training Set

$X$  คือ แอตทริบิวท์ ที่ถูกเลือกให้เป็นตัวจำแนกข้อมูล

$\text{Info}(T)$  เป็นฟังก์ชัน ที่ระบุปริมาณข้อมูลที่ต้องการเพื่อให้สามารถจำแนกประเภทที่ต้องการได้

$$\text{Info}(T) = - \sum_{j=1}^k [\text{freq}(C_j, T) / |T|] \times \log_2 [\text{freq}(C_j, T) / |T|] \quad (2.2)$$

เมื่อ  $|T|$  คือ จำนวนข้อมูลทั้งหมดใน Training Datasets

$\text{Freq}(C_j, T)$  คือ ความถี่ที่ข้อมูลใน  $T$  ปรากฏเป็นคลาส  $C_j$

$\text{Info}_x(T)$  คือ ฟังก์ชันที่ระบุปริมาณข้อมูลที่ต้องการเพื่อการจำแนกประเภทของข้อมูลโดยใช้แอตทริบิวต์  $X$  เป็นตัวตรวจสอบเพื่อแยกข้อมูล

$$\text{Info}_x(T) = \sum_{i=1}^n (|T_i| / |T|) \times \text{Info}(T_i) \quad (2.3)$$

เมื่อ  $i$  คือ จำนวนค่าที่เป็นไปได้ของแอตทริบิวต์  $x$

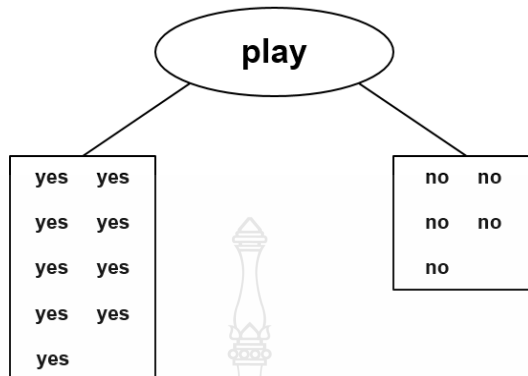
$|T_i|$  คือ จำนวนข้อมูลที่มีค่า  $x=i$

ตารางที่ 2.2 ตัวอย่างชุดข้อมูลประกอบการตัดสินใจเล่นกีฬาชนิดหนึ่ง

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
cloudy	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
cloudy	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
cloudy	mild	high	TRUE	yes
cloudy	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

ข้อมูลที่จะใช้ประกอบการอธิบายอัลกอริทึมนี้ตามตารางที่ 2.2 เป็นข้อมูลสภาพอากาศที่ใช้ประกอบการตัดสินใจการเล่นกีฬาชนิดหนึ่งว่าสภาพอากาศอย่างไรจึงจะเล่น (play = yes) และสภาพอากาศอย่างไรจึงไม่เล่น (play = no) ข้อมูลที่เป็นจุดมุ่งหมายในการจำแนกประเภทคือแอตทริบิวต์ play โดย attribute outlook, temperature, humidity, windy ทำหน้าที่เป็นแอตทริบิวต์

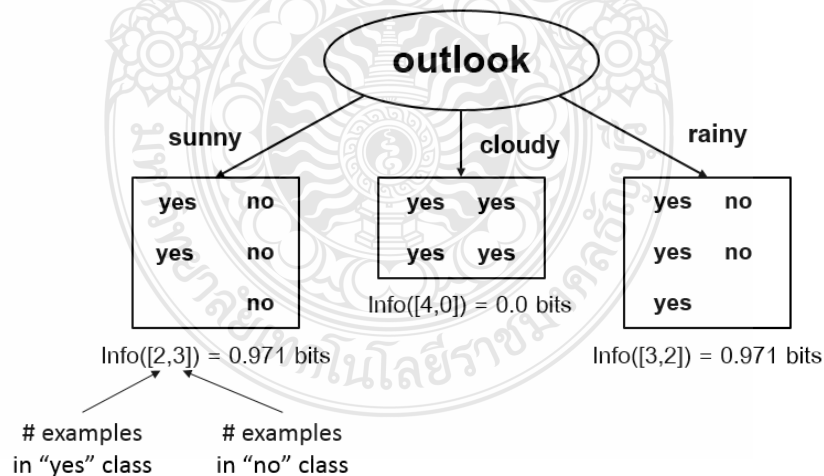
ประกอบการทำนาย จากตัวอย่างข้อมูลสภาพอากาศ เซตของข้อมูลฝึก T ประกอบด้วยข้อมูล 2 คลาส คือ play = yes และ play = no ข้อมูลประกอบการตัดสินใจจำแนกคลาสได้ดังแสดงในรูปที่ 2.10



รูปที่ 2.10 ชุดข้อมูลของ play [16]

$$\begin{aligned} \text{Info}(T) &= -(9/14) \times \log_2(9/14) - (5/14) \times \log_2(5/14) \\ &= 0.940 \text{ bits} \end{aligned}$$

การจะจำแนกคลาสของข้อมูลออกเป็น play = yes หรือ play = no ต้องใช้ข้อมูลจากแอตทริบิวต์อื่นประกอบการตัดสินใจ ถ้าเลือก attribute outlook จะต้องการปริมาณข้อมูลเพิ่มเพื่อประกอบการเลือกคลาสดังแสดงในรูปที่ 2.11



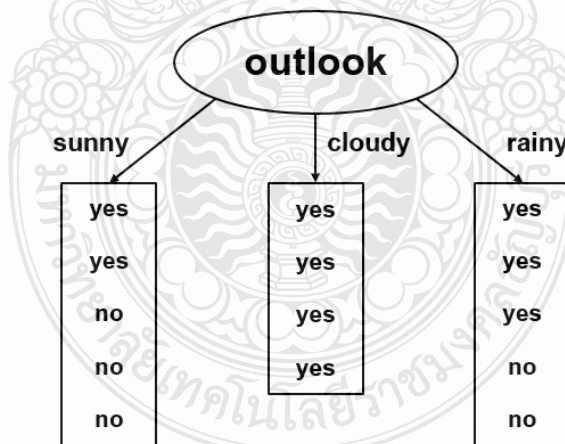
รูปที่ 2.11 ชุดข้อมูลที่นำมาใช้คำนวณของ outlook [16]

$$\begin{aligned}
\text{Info}_{\text{outlook}}(T) &= (5/14) \times [-(2/5) \times \log_2(2/5) - (3/5) \times \log_2(3/5)] \\
&\quad + (4/14) \times [-(4/4) \times \log_2(4/4) - (0/4) \times \log_2(0/4)] \\
&\quad + (5/14) \times [-(3/5) \times \log_2(3/5) - (2/5) \times \log_2(2/5)] \\
&= 0.693 \text{ bits}
\end{aligned}$$

นั่นคือ ถ้ามีข้อมูลใหม่เข้ามา เมื่อพิจารณาจากค่า outlook ของข้อมูลใหม่นี้ จะต้องใช้ข้อมูลเพิ่มอีก 0.693 bits จึงจะบอกคลาสข้อมูลที่ถูกต้องของข้อมูลใหม่นี้ได้ ค่า  $\text{Info}(T)$  เรียกได้อีกอย่างว่า ค่า entropy

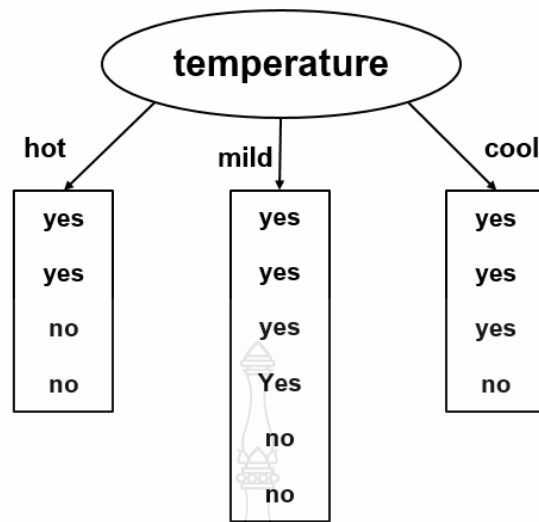
$$\begin{aligned}
\text{entropy}(P_1, P_2, \dots, P_n) &= -P_1 \log P_1 - P_2 \log P_2 - \dots - P_n \log P_n \\
\text{Info}([2,4,3]) &= \text{entropy}(2/9, 4/9, 3/9) \\
&= -(2/9) \log(2/9) - (4/9) \log(4/9) - (3/9) \log(3/9) \\
&= (-2 \log 2 - 4 \log 4 - 3 \log 3 + 9 \log 9)/9
\end{aligned}$$

แอตทริบิวต์ที่สามารถถูกเลือกมาเป็นตัวทดสอบเพื่อจัดกลุ่มของข้อมูลฝึก คือ outlook, temperature, humidity และ windy คำนวณค่า gain ได้ดังแสดงในรูปที่ 2.12 - 2.15



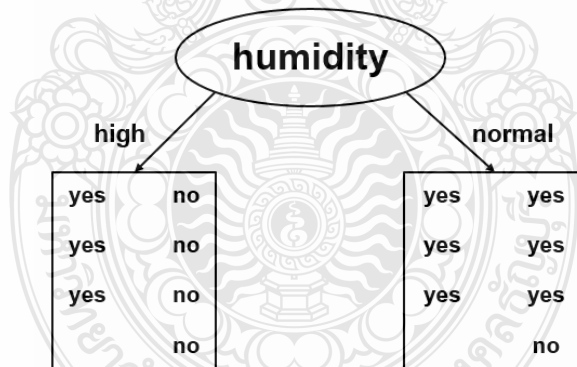
รูปที่ 2.12 ชุดข้อมูลของ outlook [16]

$$\begin{aligned}
\text{Gain}(\text{outlook}) &= \text{Info}(T) - \text{Info}_{\text{outlook}}(T) \\
&= 0.940 - 0.693 \\
&= 0.247 \text{ bits}
\end{aligned}$$



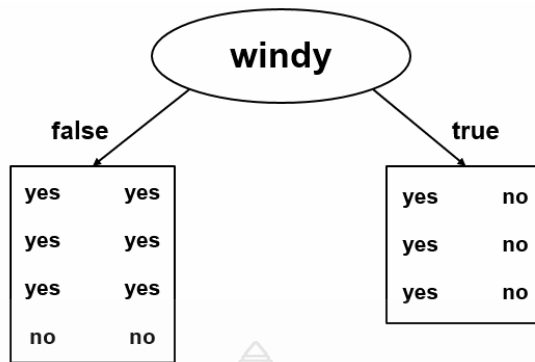
รูปที่ 2.13 ชุดข้อมูลของ temperature [16]

$$\begin{aligned}
 \text{Gain}(\text{temperature}) &= \text{Info}(T) - \text{Info}_{\text{temperature}}(T) \\
 &= 0.940 - 0.911 \\
 &= 0.029 \text{ bits}
 \end{aligned}$$



รูปที่ 2.14 ชุดข้อมูลของ humidity [16]

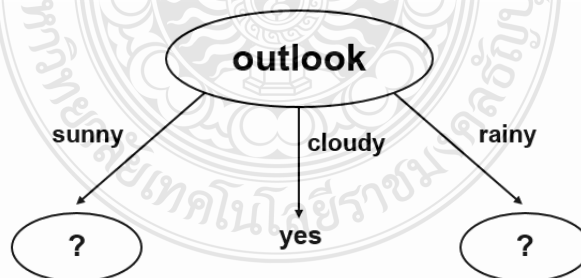
$$\begin{aligned}
 \text{Gain}(\text{humidity}) &= \text{Info}(T) - \text{Info}_{\text{humidity}}(T) \\
 &= 0.940 - 0.788 \\
 &= 0.152 \text{ bits}
 \end{aligned}$$



รูปที่ 2.15 ชุดข้อมูลของ windy [16]

$$\begin{aligned}
 \text{Gain(windy)} &= \text{Info}(T) - \text{Info}_{\text{windy}}(T) \\
 &= 0.940 - 0.892 \\
 &= 0.048 \text{ bits}
 \end{aligned}$$

แอตทริบิวต์ที่ให้ค่า gain สูงที่สุดคือ outlook ดังนั้นแอตทริบิวต์ outlook จึงถูกเลือกเป็น โหนดรากของต้นไม้ตัดสินใจ เนื่องจากแอตทริบิวต์ outlook ยังไม่สามารถจัดกลุ่มเป็นคลาสเดียวกันทั้งหมด (outlook = sunny จัดกลุ่มข้อมูลเป็นคลาส yes จำนวน 2 เรคคอร์ด และคลาส no จำนวน 3 เรคคอร์ด และ outlook = rainy จัดกลุ่มข้อมูลเป็นคลาส yes จำนวน 3 เรคคอร์ด และคลาส no จำนวน 2 เรคคอร์ด) จึงต้องสร้างต้นไม้ตัดสินใจโหนดต่อไป โดยพิจารณาเลือกแอตทริบิวต์ที่จะมาเป็นโหนดในระดับที่ 2 ต่อจากโหนดรากในกรณี outlook = cloudy ไม่จำเป็นต้องสร้างโหนดเพิ่มเติม เนื่องจากสามารถจัดกลุ่มข้อมูลที่เป็นคลาส yes ได้ทั้งหมดดังแสดงในรูปที่ 2.16



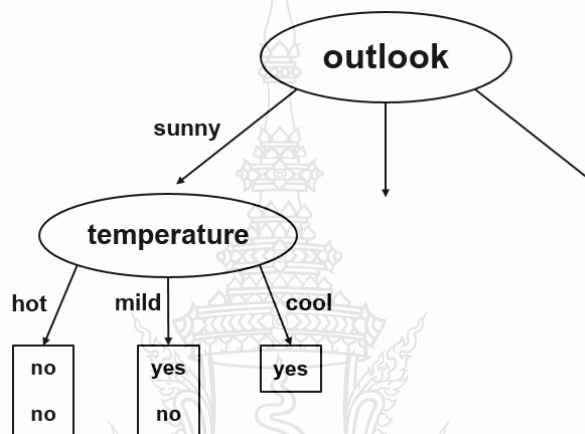
รูปที่ 2.16 ลักษณะการแยกโหนดในระดับที่ 2 ต่อจาก outlook (root node) [16]

แอตทริบิวต์ที่สามารถถูกเลือกเป็นโหนดในระดับที่ 2 ประกอบด้วย temperature, humidity และ windy (attribute outlook จะไม่ถูกใช้อีก เพราะสภาพอากาศจะไม่มีโอกาสเกิดเหตุการณ์ outlook = sunny และ outlook = rainy)

พิจารณาการสร้างโหนดลูกทางด้านซ้ายมือ (outlook = sunny) ถ้าเลือก attribute temperature จะคำนวณค่า gain ได้ดังนี้

$$\text{Gain}(\text{temperature}) = \text{Info}(\text{outlook}=\text{sunny}) - \text{Info}_{\text{temperature}}(\text{outlook}=\text{sunny})$$

เนื่องจาก outlook = sunny จัดกลุ่มข้อมูลที่เป็นคลาส yes 2 เรคคอร์ด และ ข้อมูลที่เป็นคลาส no 3 เรคคอร์ด ดังแสดงในรูปที่ 2.17



รูปที่ 2.17 ลักษณะการแยกโหนดในระดับที่ 3 (temperature) ต่อจาก sunny [16]

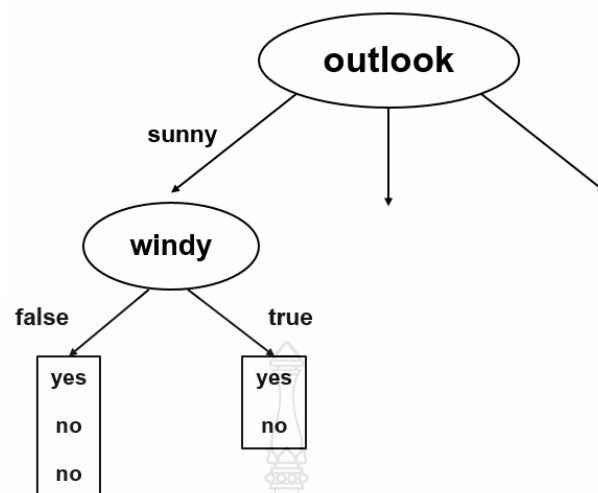
$$\begin{aligned} \text{Info}(\text{outlook}=\text{sunny}) &= -(2/5) \times \log_2(2/5) - (3/5) \times \log_2(3/5) \\ &= 0.971 \text{ bits} \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{temperature}}(\text{outlook}=\text{sunny}) &= \text{Info}([0,2], [1,1], [1,0]) \\ &= (2/5) \times [-(0/2) \times \log_2(0/2) - (2/2) \times \log_2(2/2)] \\ &\quad + (2/5) \times [-(1/2) \times \log_2(1/2) - (1/2) \times \log_2(1/2)] \\ &\quad + (1/5) \times [-(1/1) \times \log_2(1/1) - (0/1) \times \log_2(0/1)] \\ &= 0.4 \text{ bits} \end{aligned}$$

$$\begin{aligned} \therefore \text{Gain}(\text{temperature}) &= 0.971 - 0.4 \text{ bits} \\ &= 0.571 \text{ bits} \end{aligned}$$

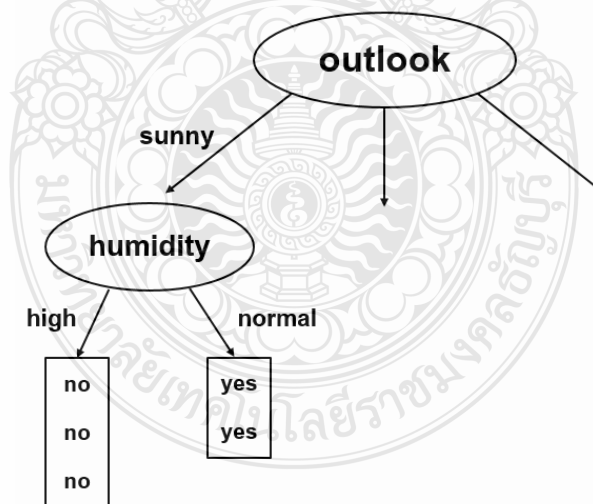
เมื่อ outlook = sunny แล้วทดลองจำแนกกลุ่มข้อมูลต่อไปด้วย attribute windy และ attribute humidity ดังแสดงในรูปที่ 2.18 - 2.19





รูปที่ 2.18 ลักษณะการแยกโหนดในระดับที่ 3 (windy) ต่อจาก sunny [16]

$$\begin{aligned}
 \text{Gain}(\text{windy}) &= \text{Info}(\text{outlook}=\text{sunny}) - \text{Info}_{\text{windy}}(\text{outlook}=\text{sunny}) \\
 &= 0.971 - \text{Info}([1,2], [1,1]) \\
 &= 0.971 - 0.951 \text{ bits} \\
 &= 0.020 \text{ bits}
 \end{aligned}$$

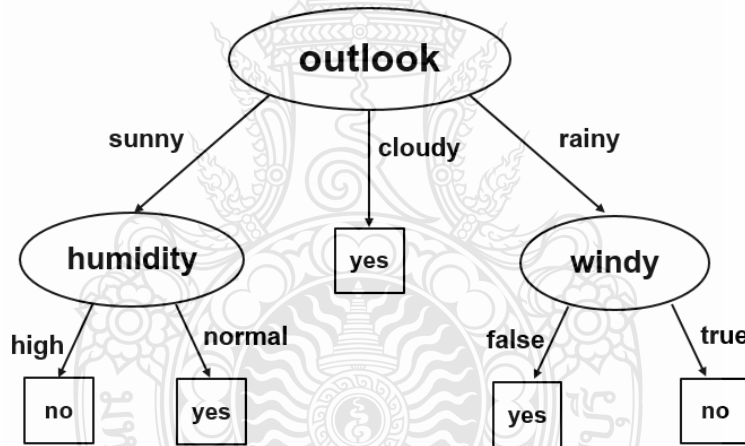


รูปที่ 2.19 ลักษณะการแยกโหนดในระดับที่ 3 (humidity) ต่อจาก sunny [16]

$$\begin{aligned}
\text{Gain}(\text{humidity}) &= \text{Info}(\text{outlook}=\text{sunny}) - \text{Info}_{\text{humidity}}(\text{outlook}=\text{sunny}) \\
&= 0.971 - \text{Info}([0,3], [2,0]) \\
&= 0.971 - 0 \text{ bits} \\
&= 0.971 \text{ bits}
\end{aligned}$$

โดยสรุปแล้ว ค่า  $\text{gain}(\text{humidity})$  มีค่ามากที่สุด จึงพิจารณาเลือก attribute humidity เป็น โหนดในระดับที่ 2 ต่อจากโหนด outlook

ต้นไม้ตัดสินใจยังเหลือ โหนดลูกทางขวาของโหนด outlook ที่ต้องพิจารณาเลือกแอตทริบิวต์และวิธีการคำนวณค่า gain ที่แสดงด้วยตัวอย่างก่อนหน้านี้ สามารถเลือกได้ว่า attribute windy จะให้ค่า gain ที่สูงที่สุด (จากกลุ่มของ attribute temperature, humidity และ windy) กระบวนการสร้าง decision tree จะสิ้นสุดเมื่อ leaf nodes เป็นกลุ่มของข้อมูลคลาสเดียวกันทั้งหมด ดังแสดงในรูปที่ 2.20



รูปที่ 2.20 ลักษณะของต้นไม้ตัดสินใจที่ได้จากข้อมูลการเล่นกีฬา [16]

ต้นไม้ตัดสินใจสามารถแปลงเป็นกฎแบบมีเงื่อนไขได้ดังนี้

- |  |                 |
|--|-----------------|
| Rule 1: IF (outlook = sunny) AND (humidity = high)   | THEN play = no  |
| Rule 2: IF (outlook = sunny) AND (humidity = normal) | THEN play = yes |
| Rule 3: IF (outlook = cloudy)                        | THEN play = yes |
| Rule 4: IF (outlook = rainy) AND (windy = false)     | THEN play = yes |
| Rule 5: IF (outlook = rainy) AND (windy = true)      | THEN play = no  |

ในกรณีที่มีข้อมูลใหม่ที่ยังไม่ทราบคลาส เช่น outlook = sunny, temperature = cool, humidity = high, windy = true สามารถใช้ต้นไม้ตัดสินใจทำนายคลาสของข้อมูลนี้ว่า play = no โดยพิจารณาจากเพียง 2 แอตทริบิวต์ คือ outlook = sunny และ humidity = high

#### 2.4.3.2 ต้นไม้การตัดสินใจแบบ C5.0

อัลกอริทึม C5.0 เป็นอัลกอริทึมที่มีการเรียนรู้แบบอาศัยการดูแลโครงสร้างผลลัพธ์ที่แทนด้วยโครงสร้างต้นไม้ Niu et al. [17] นำเสนอกระบวนการทำงานของอัลกอริทึม C5.0 โดยใช้ค่าทฤษฎีข่าวสารของคุณสมบัติที่มีค่ามากที่สุดเป็นโหนดเริ่มต้น และทำกระบวนการเดิมย้อนกลับ โดยใช้ค่า Information Entropy ร่วมด้วย เพื่อคำนวณค่าของคุณสมบัติ A เริ่มต้น หาค่า Information Entropy จากสมการที่ (2.4)

$$Info(S) = \sum_{i=1}^n P_i \log(P_i) \quad (2.4)$$

เมื่อ ค่า S หมายถึง ลำดับชุดข้อมูลตัวอย่าง และ ค่า  $P_i$  หมายถึง ค่าความน่าจะเป็นที่ได้จากการสุ่มตัวอย่างที่เกิดขึ้นใน S โดยการวัดค่า Information Gain เพื่อเป็นการสร้างลำดับตั้งสมการที่ (2.5)

$$Gain(S, A) = Info(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} \times Info(S) \quad (2.5)$$

ข้อมูลการแบ่งแยกออกของข้อมูล (Split Information) ตามสมการที่ (2.6) คือ ค่า Entropy ของแต่ละคุณสมบัติของ A โดยจะต้องทำการขจัดค่าโน้มเอียง (Bias) ของคุณสมบัติ

$$SplitInfo(S, A) = \sum_{i=1}^n \frac{|S_i|}{|S|} \log\left(\frac{|S_i|}{|S|}\right) \quad (2.6)$$

จากนั้นหาค่า Gain Ratio ด้วยการนำค่า Gain ที่คำนวณได้หารด้วยค่า Split Information ดังสมการที่ (2.7)

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (2.7)$$

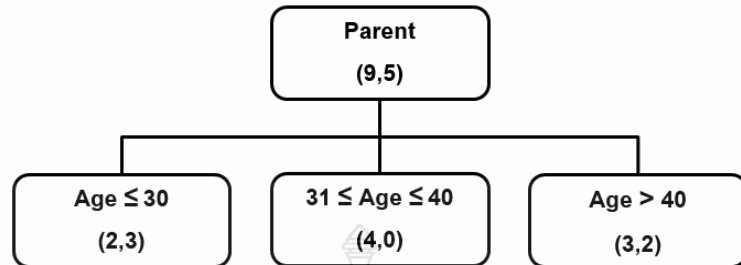
อัลกอริทึม C5.0 [18-19] พัฒนามาจากอัลกอริทึม C4.5 จึงได้เพิ่มความถูกต้องของตัวตัดแยก C5.0 ด้วยการใช้วิธี Boosting เป็นเทคนิคการสร้างการเชื่อมต่อของตัวตัดแยกหลาย ๆ ตัวให้มีความถูกต้องมากขึ้น อัลกอริทึม C5.0 ได้รวมเอาฟังก์ชันต่าง ๆ เช่น ค่าความแปรปรวนของการตัดแยกผิด (variable misclassification costs) ซึ่งจะมีการแยกค่าสูญหายในการตัดแยกของกลุ่มข้อมูล ทำให้ลดการเกิดความแปรปรวนของการตัดแยกผิดได้ ในอัลกอริทึม C5.0 ยังให้ความสำคัญในการกระจายน้ำหนักบนกลุ่มข้อมูล แต่ถ้าเกิดกรณีที่มีการตัดแยกผิดก็จะมีการเพิ่มค่าน้ำหนักให้สูงขึ้นและทำการตัดแยกข้อมูลใหม่อีกครั้ง

ข้อมูลที่จะใช้ประกอบการอธิบายอัลกอริทึมนี้ตามตารางที่ 2.3 โดยมีคลาสในการซื้อเครื่องปรับอากาศ กำหนดไว้ 2 ค่า คือ Yes, No สิ่งที่น่าสนใจคือ อายุ แบ่งออกได้ 3 ค่า คือน้อยกว่า 30, ระหว่าง 31-40 และมากกว่า 40 รายได้ แบ่งออกเป็น 3 ค่า คือ High, Medium และ Low สถานะการทำงาน แบ่งออกเป็น 2 ค่า คือ Yes, No และเครดิต แบ่งออกเป็น 2 ค่า คือ Fair, Excellent

**ตารางที่ 2.3** ตัวอย่างชุดข้อมูลการซื้อเครื่องปรับอากาศ

Age	Income	Working	Credit	Product (Air condition)
<30	High	No	Fair	No
<30	High	No	Excellent	No
31-41	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	Yes
31-41	Low	Yes	Excellent	Yes
<30	Medium	No	Fair	No
<30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<30	Medium	Yes	Excellent	Yes
31-40	Medium	No	Excellent	Yes
31-41	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

จากตารางที่ 2.3 ตัวอย่างชุดข้อมูลการซื้อเครื่องปรับอากาศ ถ้าพิจารณาที่ Age จะสามารถแบ่งกลุ่มข้อมูลได้ดังแสดงในรูปที่ 2.21



รูปที่ 2.21 ตัวอย่างการแบ่งกลุ่มข้อมูลที่ Attribute Age [18]

เมื่อนำมาคำนวณหาค่า Information ของแต่ละกลุ่มจะได้ดังนี้

$$\begin{aligned} \text{Parent} & : \quad \text{Info}([9,5]) = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14) \\ & = 0.940 \end{aligned}$$

$$\begin{aligned} \text{Age} < 30 & : \quad \text{Info}([2,3]) = - (2/5) \log_2 (2/5) - (3/5) \log_2 (3/5) \\ & = 0.971 \end{aligned}$$

$$\begin{aligned} 31 < \text{Age} < 40 & : \quad \text{Info}([4,0]) = - (4/4) \log_2 (4/4) - (0/4) \log_2 (0/4) \\ & = 0.0 \end{aligned}$$

$$\begin{aligned} \text{Age} > 40 & : \quad \text{Info}([3, 2]) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) \\ & = 0.971 \end{aligned}$$

ดังนั้นค่าของ Information เมื่อพิจารณาที่ Attribute “Age” คือ

$$\begin{aligned} \text{Info}([3,2], [4,0], [2,3]) & = [(5/14) \times 0.971] + [(4/14) \times 0] + [(5/14) \times 0.971] \\ & = 0.693 \end{aligned}$$

และค่าของ Gain เมื่อพิจารณาที่ Attribute “Age” คือ

$$\begin{aligned} \text{Gain} & = \text{Info}([9,5]) - \text{Info}([3,2], [4,0], [2,3]) \\ & = 0.940 - 0.693 \\ & = 0.247 \end{aligned}$$

Split Info ใช้หลักการเดิมในการหา Information ออกมาเป็น 3 กลุ่ม โดยการนับจำนวนรายการของข้อมูล ไม่ได้นับจำนวน ค่า Yes หรือ No จากข้อมูลดังกล่าวจะทราบว่า

กลุ่ม อายุน้อยกว่าหรือเท่ากับ 30 มีจำนวน 5 รายการ

กลุ่ม อายุน้อยกว่าหรือเท่ากับ 30 - 40 มีจำนวน 4 รายการ

กลุ่ม อายุน้อยกว่าหรือเท่ากับ 40 มีจำนวน 5 รายการ

จะได้ค่า Split Info คือ

$$\begin{aligned} \text{Split Info}([5,4,5]) &= - (5/14) \log_2 (5/14) - (4/14) \log_2 (4/14) - (5/14) \log_2 (5/14) \\ &= 1.577 \end{aligned}$$

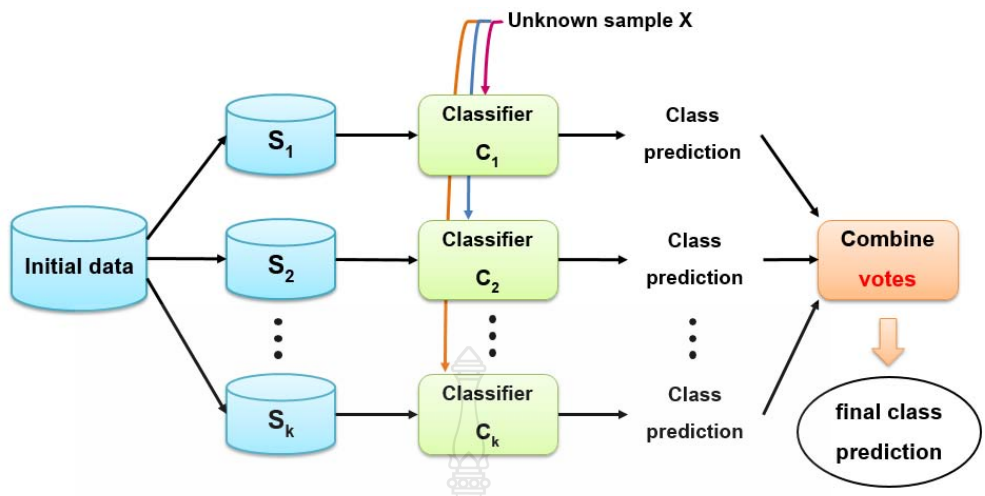
$$\begin{aligned} \therefore \text{Gain ratio} &= 0.247/1.577 \\ &= 0.156 \end{aligned}$$

เมื่อได้ค่า Gain ratio และทำการคำนวณในรายการอื่น ๆ จนครบ ก็จะสามารถแสดงได้ว่าเราจะแบ่งการทำงานอยู่ค่าของรายการใด

## 2.5 เทคนิคการเพิ่มประสิทธิภาพการคัดแยกข้อมูล

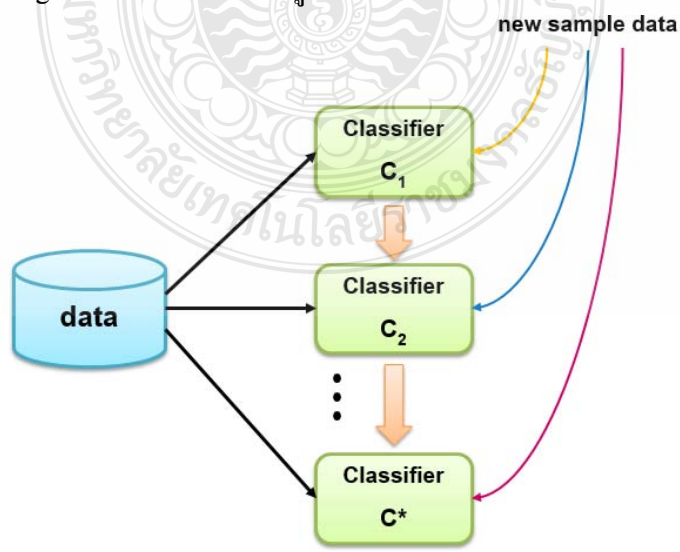
การเพิ่มประสิทธิภาพการคัดแยกข้อมูล [20] เป็นการเพิ่มความถูกต้องในการทำนายข้อมูลซึ่งทำได้ด้วยการวิธี classification หลาย ๆ ครั้ง หรือเรียกว่า multiple learning เทคนิคที่นิยมใช้ ได้แก่ Bagging และ Boosting [21]

Bagging หรือ Bootstrap aggregation เป็นการทำการคัดแยกข้อมูลหลายครั้ง เช่น 10 ครั้ง ทำให้ได้ตัวคัดแยกข้อมูล 10 โมเดล เมื่อนำทั้ง 10 โมเดล ไปใช้เพื่อการทำนายหรือจำแนกประเภทข้อมูลใหม่ จะส่งข้อมูลใหม่รายการนั้นไปให้ทั้ง 10 โมเดล เพื่อการทำนายผล ซึ่งผลที่ได้จากการทำนายของทั้ง 10 โมเดลอาจจะเหมือนกันทั้งหมด หรืออาจจะมีการทำนายผลที่ได้แตกต่างกันไป ผลลัพธ์สุดท้ายของการทำนายจะใช้วิธีนับโหวตว่าทั้ง 10 โมเดลนั้น ส่วนใหญ่ทำนายว่าเป็นผลลัพธ์ส่วนใด จะแสดงผลลัพธ์ที่ได้เป็นการทำนายของโมเดลส่วนใหญ่ วิธีการ Bagging มีขั้นตอนดังแสดงในรูปที่ 2.22



รูปที่ 2.22 เทคนิค Bagging เพื่อเพิ่มประสิทธิภาพการจำแนกข้อมูล [20]

Boosting เป็นการทำการคัดแยกข้อมูลหลายครั้งเหมือนกับเทคนิค Bagging แต่ต่างกันตรงที่การสร้างตัวคัดแยกข้อมูลแต่ละครั้งไม่ได้เป็นอิสระต่อกัน เทคนิค Boosting จะเริ่มต้นทำงานด้วยการให้ค่าน้ำหนักกับข้อมูลแต่ละเรคคอร์ด โดยเริ่มต้นทุกเรคคอร์ดมีน้ำหนักเป็น 1 เท่ากัน เมื่อทำการแยกข้อมูลในครั้งแรกและทดสอบผลการแยกข้อมูล เรคคอร์ดใดที่ตัวแยกข้อมูลทำนายผิดจะถูกเพิ่มค่าน้ำหนักให้กับการทำการคัดแยกข้อมูลครั้งต่อไปเพื่อให้ความสนใจกับข้อมูลที่ถูกทำนายผิดมากกว่าข้อมูลอื่น การทำการคัดแยกข้อมูลจะดำเนินไปเช่นนี้หลายครั้งจนกระทั่งได้ผลลัพธ์ที่มีความถูกต้องแม่นยำสูงถึงเกณฑ์ที่กำหนด หรือต่อเมื่อไม่สามารถเพิ่มความแม่นยำได้อีกต่อไป กระบวนการจึงสิ้นสุด และตัวคัดแยกข้อมูลล่าสุดที่ได้จะถูกนำไปใช้ในการทำนายเพื่อคัดแยกข้อมูลที่จะเกิดขึ้นในอนาคต วิธีการ Boosting มีขั้นตอนดังแสดงในรูปที่ 2.23



รูปที่ 2.23 เทคนิค Boosting เพื่อเพิ่มประสิทธิภาพการจำแนกข้อมูล [20]

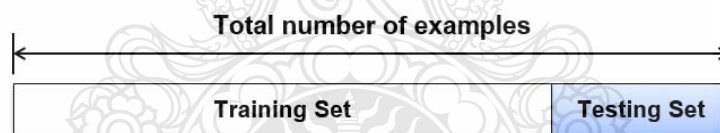
## 2.6 เทคนิคการวัดประสิทธิภาพ

ในการทดลองจะต้องมีการคำนวณหาค่าความถูกต้องของข้อมูล (Accuracy) เพื่อวัดว่าผลที่ได้จากการทำนายมีค่าความถูกต้องมากน้อยเพียงใดซึ่งคิดเป็นร้อยละ [22] ดังสมการที่ (2.8)

$$\text{ค่าความถูกต้อง (\%)} = \frac{\text{จำนวนข้อมูลที่ทำนายถูกต้อง}}{\text{จำนวนข้อมูลทั้งหมด}} \times 100 \quad (2.8)$$

ในการวัดประสิทธิภาพโดยใช้ค่าความถูกต้องของข้อมูล (Accuracy) จะต้องทำการเลือกข้อมูลชุดสอน (Training Set) และข้อมูลชุดทดสอบ (Testing Set) ซึ่งมีหลักการเลือก 3 วิธีคือเลือกสุ่มข้อมูลแบบร้อยละ เลือกสุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม (K-Fold Cross Validation) และเลือกสุ่มแบบข้อมูลแบบ Leave-one-out Cross Validation ซึ่งมีรายละเอียดดังนี้ [23]

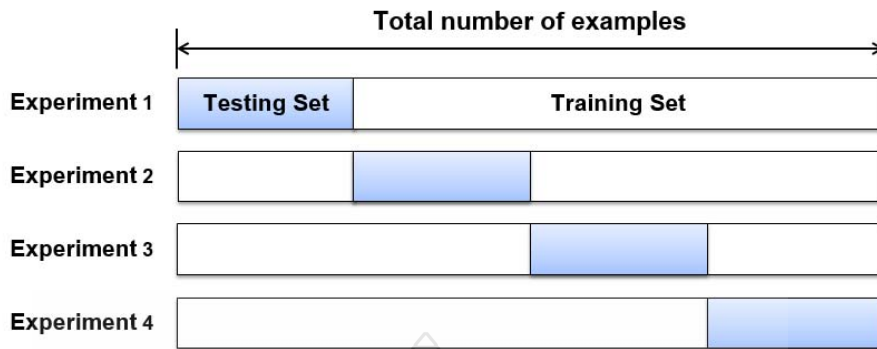
2.6.1 การเลือกสุ่มข้อมูลแบบร้อยละ (Percentage) จะเลือกสุ่มข้อมูลชุดสอนตามร้อยละที่กำหนด สำหรับข้อมูลที่เหลือจะเป็นข้อมูลชุดทดสอบดังแสดงในรูปที่ 2.24 ข้อดีของการเลือกสุ่มข้อมูลแบบร้อยละคือเป็นวิธีการเลือกสุ่มข้อมูลที่ง่าย แต่ข้อเสียคือข้อมูลทุกตัวไม่ได้ถูกนำมาเป็นข้อมูลชุดสอนและชุดทดสอบ



รูปที่ 2.24 การเลือกสุ่มข้อมูลแบบร้อยละ [23]

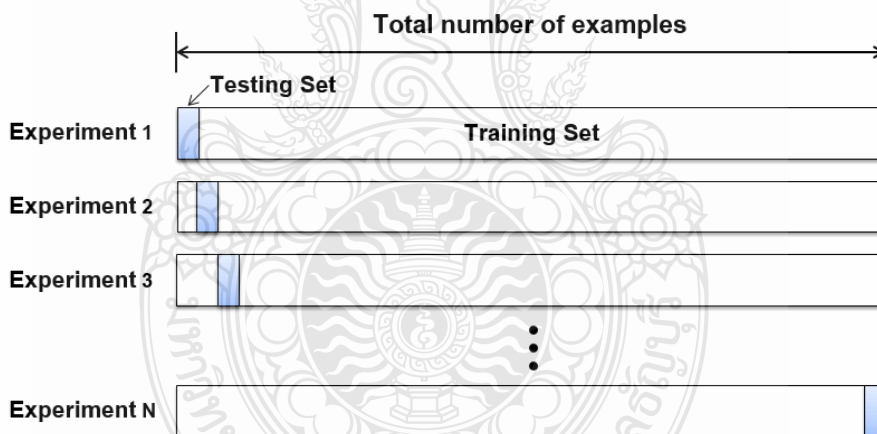
2.6.2 การเลือกสุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม (K-Fold Cross Validation) จะเลือกสุ่มข้อมูลออกเป็น K ชุดเท่ากัน ในการทดลองครั้งแรกข้อมูลชุดที่ 1 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ในการทดลองครั้งที่สองข้อมูลชุดที่ 2 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำจนกระทั่งข้อมูลทุกชุดได้ถูกนำมาเป็นข้อมูลชุดทดสอบซึ่งมีการทดลองทั้งหมด K ครั้ง ซึ่งผลลัพธ์ที่ได้จะมาจากค่าเฉลี่ยความถูกต้องของการจำแนกประเภทในแต่ละรอบ ตัวอย่างการเลือกสุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม เมื่อ  $K = 4$  ดังแสดงในรูปที่ 2.25 ข้อดีของการเลือกสุ่มข้อมูลแบบความเที่ยงตรง K กลุ่มคือ ข้อมูลทุกตัวจะถูกนำมาเป็นข้อมูลชุดสอนและข้อมูลชุดทดสอบ แต่ข้อเสียคือใช้เวลานานในการทดลอง เนื่องจากต้องทดลองข้อมูลทั้งหมด K ครั้ง ทั้งนี้นิยมกำหนดให้ค่า K มีค่าเท่ากับ 10





รูปที่ 2.25 การเลือกกลุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม เมื่อ  $K = 4$  [23]

2.6.3 เลือกกลุ่มข้อมูลแบบ Leave-one-out Cross Validation คือการเลือกกลุ่มข้อมูลแบบความเที่ยงตรง K กลุ่ม เมื่อกำหนดให้ K มีค่าเท่ากับจำนวนแถวข้อมูลทั้งหมด (N) ดังแสดงในรูปที่ 2.26 ข้อดีของการเลือกกลุ่มข้อมูลแบบ Leave-one-out Cross Validation คือเหมาะสำหรับข้อมูลขนาดเล็ก แต่ข้อเสียคือไม่เหมาะสำหรับข้อมูลขนาดใหญ่เนื่องจากต้องทดสอบหลายครั้งทำให้ใช้เวลาในการทดลองนาน



รูปที่ 2.26 การเลือกกลุ่มข้อมูลแบบ Leave-one-out Cross Validation [23]

หลังจากเลือกเทคนิคการวัดประสิทธิภาพแล้วจะต้องนำข้อมูลที่เกี่ยวข้องกับการแบ่งแยกข้อมูลจริง กับข้อมูลที่เกิดจากการทำนาย ด้วยระบบการแบ่งแยก (Classification System) เช่น การหาค่าข้อมูลแบบขรรคมดาในเมทริกซ์ (Confusion Matrix) [24] เป็นข้อมูลสำคัญที่จะใช้วัดความแม่นยำของโมเดลดังตารางที่ 2.4 เป็นการทำนายข้อมูลที่มีสองคลาส คือ คลาส positive และ คลาส negative

ตารางที่ 2.4 ตัวอย่างระบบการแบ่งแยกแบบ Confusion Matrix

		ค่าทำนาย (Predicted)	
		ปฏิเสธ (Negative)	ยอมรับ (Positive)
ค่าความจริง (Actual)	ปฏิเสธ (Negative)	A	B
	ยอมรับ (Positive)	C	D

จากตาราง Confusion Matrix มีความหมายดังนี้

ค่า A เป็นตัวเลขที่ถูกต้องจากการทำนาย

ค่า B เป็นตัวเลขที่ไม่ถูกต้องจากการทำนาย

ค่า C เป็นตัวเลขที่ไม่ถูกต้องจากการทำนาย

ค่า D เป็นตัวเลขที่ถูกต้องจากการทำนาย

โมเดลที่สร้างขึ้นจะต้องมีการวัดเจาะจงว่า การทำนายถูกต้องนั้นเป็น True positive rate เท่าใด และเป็น True negative rate เท่าใด ซึ่งจะตัดสินได้ว่าผลการทำนายเที่ยงตรงเพียงใด (Precision) สุดท้ายจึงได้ค่าความแม่นยำตรง (Accuracy) ของ โมเดล ตัววัดเหล่านี้สามารถคำนวณค่าได้จาก confusion matrix ดังนี้

$$\text{True positive rate (or sensitivity)} = D/(D+C)$$

$$\text{True negative rate (or specificity)} = A/(A+B)$$

$$\text{Precision} = D/(D+B)$$

$$\text{Accuracy} = (D+A)/(\text{all samples})$$

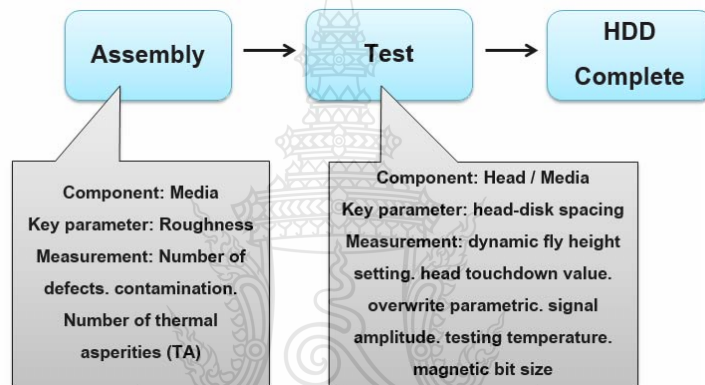
## 2.7 การสำรวจเอกสาร (Literature Reviews)

วิทยานิพนธ์นี้ได้ทำการศึกษาเกี่ยวข้องกับการตรวจสอบคุณภาพฮาร์ดดิสก์ จากการศึกษาวิธีการตรวจสอบในงานวิจัย และ วารสาร แตกต่างกัน แต่ผลลัพธ์ที่ได้คือ เพื่อเป็นแนวทางในการพัฒนา และปรับปรุงแก้ไขกระบวนการผลิตให้ดีขึ้นทั้งสิ้น สรุปได้พอสังเขปดังนี้

### 2.7.1 Customer Failure Modes Prediction for Hard Disk Drives using Neural Networks Rank-Level Fusion

Tepin W และคณะ [25] งานวิจัยนี้ได้กล่าวถึงปัญหาที่เกิดขึ้นเนื่องจากการลดระยะห่างระหว่างหัวอ่านกับจานแม่เหล็กหรือเรียกว่า HDI (Head Disk Interaction) โดยได้มีการตรวจสอบเบื้องต้นจากการกระบวนการผลิตในส่วนของการประกอบ และในส่วนของการทดสอบหลังจากการ

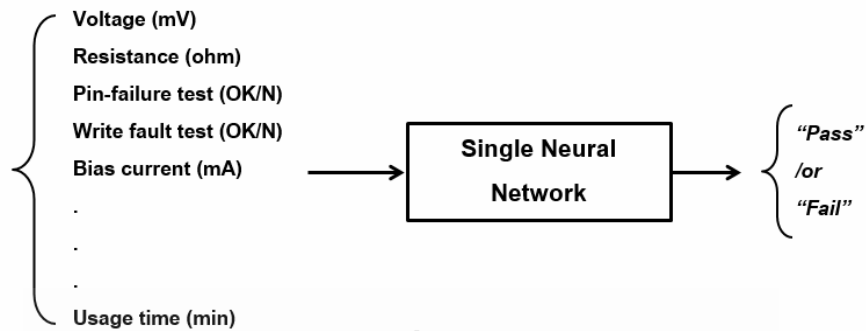
ประกอบดังแสดงในรูปที่ 2.27 ทั้งนี้เนื่องจากปัญหาในการผลิตส่วนใหญ่เกิดจากปัญหาจาก HDI (Head Disk Interaction) จำนวนมาก ในงานวิจัยนี้เริ่มจากการปรับแต่งข้อมูลจากตัวอย่างที่มีปัญหา และตัวอย่างที่ไม่มีปัญหาซึ่งได้จากการจัดกลุ่มด้วยเทคนิค Principal Component Analysis (PCA) ซึ่งสามารถช่วยลดค่า Cross-Relation จากการเลือกเวกเตอร์หลักเฉพาะ (Eigenvalue) จึงทำให้ความถูกต้องเพิ่มขึ้นประมาณ 2-3% ในส่วนของการแยกประเภทนั้นจะใช้ Neural Network (NN), Discrimination Analysis (DA), Bayesian Network (BN) and Support Vector Machines (SVM) โดยพิจารณาจาก 2 คลาสเท่านั้น ได้แก่ “ดี” และ “เสีย” โดยข้อมูลนำเข้าจะมาจาก PCA ที่ได้จากการปรับแต่งข้อมูลแล้ว ซึ่งงานวิจัยนี้ได้ทำการตั้งค่า NN ให้มีการใช้ Pruning Method โดยเริ่มจากการใช้โครงข่ายประสาทเทียมขนาดใหญ่และลดขนาดของส่วนที่ไม่จำเป็นลง



รูปที่ 2.27 การตรวจสอบเบื้องต้นในกระบวนการผลิต [25]

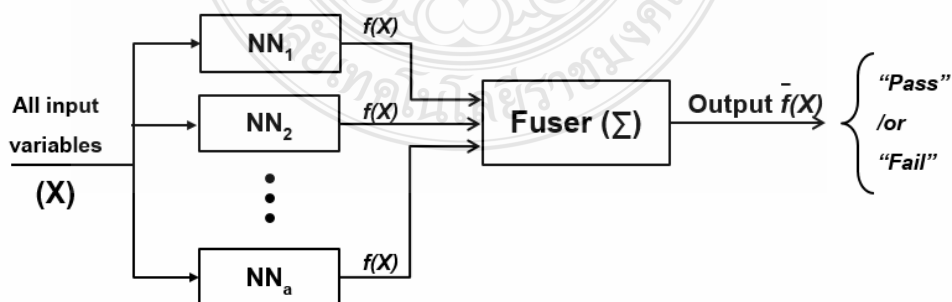
### 2.7.2 The Classification Model for Hard Disk Drive Functional Tests under Sparse Data Conditions

Chetchotsak D และคณะ [26] งานวิจัยนี้ได้นำเสนอเทคนิคโครงข่ายประสาทเทียมแบบคณะกรรมการในการตรวจสอบฮาร์ดดิสก์ในกระบวนการผลิต เพื่อระบุได้ว่าไคร์ฟที่กำลังทดสอบนั้นมีความเป็นไปได้ว่าจะ “เสีย” หรือ “ดี” โดยปกติจะใช้โครงข่ายประสาทเทียมแบบเดียวในการทำนายดังแสดงในรูปที่ 2.28



รูปที่ 2.28 แบบจำลองที่ใช้โครงข่ายประสาทเทียมแบบเดี่ยวในการทำนาย [26]

งานวิจัยนี้ได้ใช้เทคนิคการเลือกข้อมูล (Feature Selection) ซึ่งสามารถแบ่งออกเป็น 3 ขั้นตอนได้แก่ การกรองข้อมูล (Screening), การจัดลำดับข้อมูล (Ranking) และการเลือกข้อมูล (Selecting) ในขั้นตอนการกรองข้อมูลนั้นจะทำการตรวจสอบว่าตัวแปรใดที่สำคัญต่อการวิเคราะห์บ้าง โดยจะตัดข้อมูลบางประเภทได้แก่ ข้อมูลที่สูญหาย (Missing Values), ข้อมูลที่มีค่ามาก ๆ หรือ หลากหลายค่าและตัวแปรที่มีค่าคงที่อยู่ตลอดเวลา เป็นต้น จากนั้นเมื่อเข้าสู่กระบวนการการจัดลำดับข้อมูล ข้อมูลนำเข้าจะทำนายตำแหน่งที่อยู่ในการจัดอันดับของตัวแปรโดยใช้ค่า P-Values ทางสถิติเข้ามาช่วย จากนั้นอัลกอริทึมจะทำการคัดเลือกตัวแปรหลักที่ได้จากการจัดลำดับมาใช้ในการวิเคราะห์ต่อไปในส่วนของการคัดแยกนั้นจะใช้โครงข่ายประสาทเทียมเข้ามาช่วยแบ่งกลุ่มของข้อมูลระหว่าง “เสีย” และ “ดี” โดยใช้ Multilayer Perceptron แบบ Back propagation จากนั้นจะนำมาสร้างเป็นแบบจำลองการคัดแยกโดยใช้โครงข่ายประสาทเทียมแบบคณะกรรมการมาช่วย ซึ่งประกอบด้วยโครงข่ายประสาทเทียมหลายโครงข่าย ที่สามารถเรียนรู้ข้อมูลต่างชนิดกันได้โดยใช้อัลกอริทึม Bootstrap ในการสุ่มคัดเลือกข้อมูลที่จะเข้าสู่โครงข่ายประสาทเทียม จากนั้นจะนำข้อมูลที่ได้จากแต่ละโครงข่ายมารวมกันโดยใช้เทคนิค Voting ดังแสดงในรูปที่ 2.29



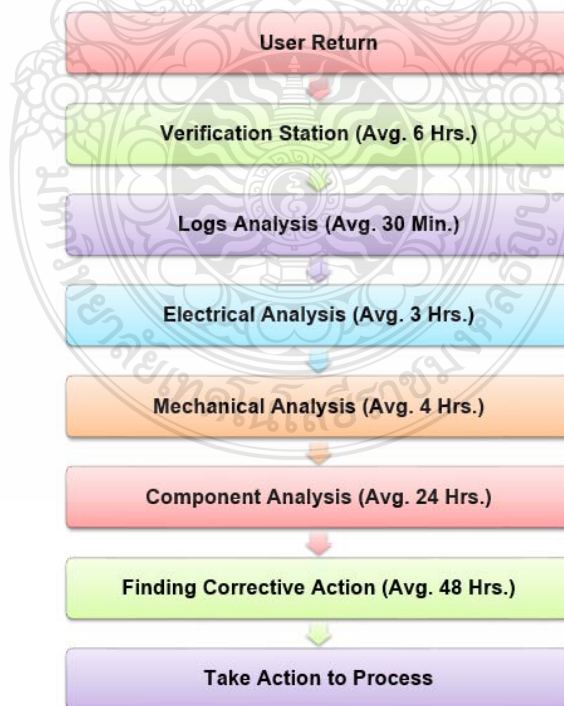
รูปที่ 2.29 แบบจำลองโมเดลของ คณะกรรมการ Network [26]

### 2.7.3 Hard Disk Drive Failure Mode Prediction from SMART Attribute using Data

#### Mining Method

Ramangkul T และคณะ [27] งานวิจัยนี้ผู้วิจัยเริ่มจากการถอดรหัสข้อมูลของสมาร์ต หรือ S.M.A.R.T (Self-Monitoring, Analysis and Reporting Technology) ให้อยู่ในรูปของแอตทริบิวต์ จากนั้นจะนำข้อมูลที่ได้ออกมาวิเคราะห์ข้อมูล ได้แก่ การเลือกแอตทริบิวต์ที่สนใจและการปรับแต่งข้อมูล เมื่อได้ข้อมูลที่พร้อมสำหรับการประมวลผลแล้วจะทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม ได้แก่ กลุ่มข้อมูลสำหรับเรียนรู้และกลุ่มข้อมูลสำหรับทดสอบ โดยใช้เทคนิคการสร้างชุดข้อมูลด้วยการแบ่งกลุ่มจากนั้นจะเข้าสู่กระบวนการสร้างแบบจำลองด้วยอัลกอริทึม C5.0 โดยใช้เทคนิค Boosting และต้นไม้ตัดสินใจแบบคณะกรรมการ (Committee Decision Tree) ซึ่งทำให้ค่าความถูกต้องของการคัดแยกสูงขึ้น

ในการทำงานจริงของผู้เชี่ยวชาญนั้นเมื่อได้รับแจ้งจากผู้ใช้จะถูกนำเข้าสู่กระบวนการทดสอบดังแสดงในรูปที่ 2.30 ผู้เชี่ยวชาญจะทำการดาวน์โหลดไฟล์สมาร์ตจากฮาร์ดดิสก์ ซึ่งอยู่ในรูปแบบไบนารี (Binary) มาทำการแปลงให้อยู่ในรูปแบบแอตทริบิวต์ จากนั้นจะทำการวิเคราะห์และหาความสัมพันธ์ระหว่างค่าที่ผิดปกติกับเวลาที่เกิดเหตุการณ์ขึ้น เพื่อสร้างสมมติฐานและออกแบบการทดสอบแจ้งจากผู้ใช้เหล่านั้นได้ โดยปกติแล้วแต่ละแอตทริบิวต์ จะสามารถบอกประเภทหรือสาเหตุการเสียหายเบื้องต้นได้



รูปที่ 2.30 กระบวนการวิเคราะห์สาเหตุการเสียหายสู่กระบวนการปรับปรุงคุณภาพการผลิต [27]

## บทที่ 3

### วิธีดำเนินการวิจัย

#### 3.1 ขั้นตอนการดำเนินงานวิจัย

3.1.1 รวบรวมข้อมูลที่ใช้สำหรับการวิเคราะห์สาเหตุการเสียชีวิตของฮาร์ดดิสก์

3.1.2 นำข้อมูลที่ได้เข้าสู่กระบวนการเตรียมข้อมูล เช่น การเลือกแอตทริบิวต์ที่สนใจและการปรับแต่งข้อมูล

3.1.3 เมื่อได้ข้อมูลที่พร้อมสำหรับการประมวลผลแล้วจะทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม ได้แก่ กลุ่มข้อมูลสำหรับเรียนรู้และกลุ่มข้อมูลสำหรับทดสอบ โดยใช้เทคนิคการสร้างชุดข้อมูลด้วยการแบ่งกลุ่มจากนั้นจะเข้าสู่กระบวนการสร้างแบบจำลองด้วยอัลกอริทึม C5.0

3.1.4 ปรับค่าสถานะต่าง ๆ และทดสอบผลการทำงานของอัลกอริทึม C5.0 โดยเปรียบเทียบกับอัลกอริทึมอื่น เช่น Neural Network, C&R Tree, SVM และ CHAID และทดสอบการทำงานของอัลกอริทึม C5.0 กับวิธีการวิเคราะห์ที่ใช้อยู่ในปัจจุบัน ซึ่งจะทดสอบโดยวิศวกรผู้เชี่ยวชาญ

3.1.5 สรุปผล และจัดทำวิทยานิพนธ์ฉบับสมบูรณ์

#### 3.2 การเตรียมข้อมูล

งานวิจัยนี้ได้นำข้อมูลที่ได้จากการทดสอบฮาร์ดดิสก์ที่ผ่านการวิเคราะห์หาสาเหตุการเสียชีวิตโดยผู้เชี่ยวชาญมาแล้วย้อนหลังเป็นเวลา 1 ปี โดยเลือกเฉพาะฮาร์ดดิสก์ประเภท ES (Enterprise Storage) และ NL (Near-Line Storage) งานวิจัยนี้ได้ทำการรวบรวมข้อมูลแอตทริบิวต์ของฮาร์ดดิสก์จำนวน 1,161 ไดรฟ์ที่มีปัญหาในกระบวนการผลิตและฮาร์ดดิสก์ที่ไม่มีปัญหาจากการผลิต ซึ่งสามารถแบ่งประเภทการเสียชีวิตได้เป็น 10 ประเภทดังตารางที่ 3.1 ในการเตรียมข้อมูลสำหรับการทดลองจะประกอบด้วยการเลือกข้อมูลซึ่งเป็นการเลือกข้อมูลแอตทริบิวต์ที่ได้จากกระบวนการทดสอบ และการปรับแต่งข้อมูลซึ่งเป็นการปรับข้อมูลที่ได้จากการเลือกข้อมูลให้สอดคล้องกับค่าที่ถูกกำหนดไว้ โดยผู้พัฒนาและเพื่อสามารถใช้งานได้เหมาะสมกับอัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID โดยมีรายละเอียดดังนี้

ตารางที่ 3.1 ประเภทการเสียทั้งหมดที่พบย้อนหลังเป็นเวลา 1 ปี

ประเภทการเสีย	จำนวน
ATI/STE	9
CND	114
Degraded Head	28
FHM	67
Head Instability	46
NMD	330
NPF	439
OTW	16
VDW	90
Weak write	22
รวม	1,161

### 3.2.1 การเลือกข้อมูลสำหรับการทดลอง

การเลือกชุดข้อมูลสำหรับวิเคราะห์นั้น จะอ้างอิงจากแอดทริบิวต์ที่ได้จากการรวบรวมข้อมูลและเป็นที่ยอมรับจากวิศวกรผู้เชี่ยวชาญและแอดทริบิวต์อื่น ๆ ที่มีความน่าจะเป็นต่อการระบุสาเหตุการเสียของฮาร์ดดิสก์ เนื่องจากข้อมูลที่ได้จากการทดสอบคุณภาพฮาร์ดดิสก์ประกอบด้วยแอดทริบิวต์จำนวนมาก และมีความสำคัญแตกต่างกัน ดังนั้นจึงจำเป็นต้องอาศัยการเลือกแอดทริบิวต์โดยวิศวกรผู้เชี่ยวชาญในแต่ละส่วนของการวิเคราะห์สาเหตุการเสีย ทั้งนี้แอดทริบิวต์ที่ใช้ในการทดลองจะได้จากผู้เชี่ยวชาญในแต่ละส่วนรวมถึงค่าที่ถูกกำหนดจำนวน 76 แอดทริบิวต์ ซึ่งถูกนำมาใช้เพื่อทำการหาความถูกต้องเหมาะสมที่สุดในแต่ละอัลกอริทึม โดยเมื่อนำข้อมูลไปทดสอบกับแต่ละอัลกอริทึมแล้วทำให้ได้ค่าผลลัพธ์ที่ถูกต้องที่สุด โดยมีรายละเอียดดังในตารางที่ 3.2

ตารางที่ 3.2 รายละเอียดการแปรค่าแอตทริบิวต์ที่นำมาทดลอง

แอตทริบิวต์	การทดสอบ	รายละเอียด
1-3	Bit Error Rate by Zone	ทดสอบจำนวนบิตที่เกิดการส่งผิดพลาดเทียบกับจำนวนที่ส่งทั้งหมด (Bit Error Rate) ในแต่ละโซนข้อมูล
4-7	Contact Detect	วัดระยะห่างระหว่างหัวอ่านและจานแม่เหล็ก
8-11	Head Resistance Test	ทดสอบค่าความต้านทานของหัวอ่านข้อมูล
12-13	Head Stability	ตรวจสอบความไม่เสถียรของหัวอ่าน โดยการอ่านเขียนข้อมูลเป็นเวลานานหลาย ๆ ครั้ง
14	Hot Cold Write Test	ทดสอบการอ่านเขียนข้อมูลในอุณหภูมิสูงและต่ำ
15	Read Primary List	ตรวจสอบตำแหน่งที่ไม่สามารถอ่านเขียนข้อมูลได้
16-20	Resonance	ตรวจสอบความถี่เรโซแนนซ์และขนาดของสัญญาณบนหัวอ่าน
21-22	Servo Actuator Retract	ทดสอบค่ากระแสที่ใช้ในการเคลื่อนตำแหน่งหัวอ่านไปยังตำแหน่งข้อมูล
23-26	Servo Flaw Scan	ค้นหาข้อบกพร่องของสัญญาณเซอร์โว
27-28	Servo Seek	การทดสอบระยะเวลาในการเข้าถึงข้อมูล
29-42	Side Erasure Test	ทดสอบการเขียนข้อมูลแล้วอ่านค่าบริเวณแทรคข้างเคียงเพื่อป้องกันผลกระทบจากการรบกวนของสนามแม่เหล็กขณะทำการเขียนข้อมูล
43-54	Skip Write Detect	ตรวจจับการเขียนข้อมูลเพื่อวัดและปรับความสูงของหัวอ่านขณะบินอยู่ที่จะส่งผลในการทำงานที่ผิดพลาด
55-70	SMART Attribute Check	ตรวจสอบค่าแอตทริบิวต์สะสมต่าง ๆ ที่อยู่ใน SMART ให้อยู่ในสเป็ค เช่น จำนวนครั้งการใช้งาน อุณหภูมิ ค่ากระแสที่ใช้ควบคุมหัวอ่าน ค่าความเสถียรของสัญญาณ จำนวนบิตที่เกิดการส่งผิดพลาด จำนวนการเขียนข้อมูลที่ผิดพลาด จำนวนการอ่านข้อมูล จำนวนการเขียนข้อมูล
71-73	Thermal Asperity Scan	ตรวจสอบหาตำแหน่งที่ผิดปกติบนแผ่นจานแม่เหล็ก ที่มีลักษณะเป็นรอยลึก หรือนูนขึ้น โดยดูจากสัญญาณ
74-76	Transfer Rate Test	ทดสอบอัตราการส่งผ่านข้อมูล



เนื่องจากข้อมูลจากระบวนการทดสอบฮาร์ดดิสก์ประกอบด้วยแอตทริบิวต์จำนวนมาก และมีความสำคัญแตกต่างกัน ดังนั้นจึงจำเป็นต้องอาศัยการเลือกแอตทริบิวต์โดยผู้เชี่ยวชาญในแต่ละส่วนของการวิเคราะห์สาเหตุการเสีย ทั้งนี้แอตทริบิวต์ที่ใช้ในการทดลองจะได้จากผู้เชี่ยวชาญในแต่ละส่วนรวมถึงค่าที่ถูกกำหนดจำนวน 76 แอตทริบิวต์ ซึ่งสามารถจัดลำดับความสำคัญได้ดังตารางที่ 3.3 โดยกลุ่มที่ 1 จะให้ความสำคัญในการวิเคราะห์หาสาเหตุการเสียของฮาร์ดดิสก์มากที่สุด รองลงมาเป็นกลุ่มที่ 2 กลุ่มที่ 3 กลุ่มที่ 4 และกลุ่มที่ 5 ตามลำดับ

**ตารางที่ 3.3** การจัดกลุ่มความสำคัญของแอตทริบิวต์โดยผู้เชี่ยวชาญ

ระดับความสำคัญของแอตทริบิวต์				
1	2	3	4	5
Attribute08	Attribute04	Attribute06	Attribute01	Attribute10
Attribute09	Attribute05	Attribute07	Attribute02	Attribute11
Attribute16	Attribute12	Attribute14	Attribute03	Attribute15
Attribute17	Attribute13	Attribute23	Attribute63	Attribute29
Attribute18	Attribute27	Attribute24	Attribute64	Attribute31
Attribute19	Attribute28	Attribute25	Attribute65	Attribute37
Attribute20	Attribute30	Attribute26	Attribute66	Attribute46
Attribute21	Attribute33	Attribute36	Attribute67	Attribute47
Attribute22	Attribute34	Attribute42	Attribute68	Attribute48
Attribute32	Attribute38	Attribute55	Attribute69	Attribute49
Attribute35	Attribute39	Attribute56	Attribute70	Attribute50
Attribute41	Attribute40	Attribute57		Attribute51
Attribute43	Attribute74	Attribute58		Attribute52
Attribute44	Attribute75	Attribute59		Attribute53
Attribute45	Attribute76	Attribute60		Attribute54
		Attribute61		Attribute71
		Attribute62		Attribute72
				Attribute73

ข้อมูลที่น่ามาประมวลผลต้องมีรูปแบบที่ถูกต้องตรงตามรูปแบบของโปรแกรมที่เป็น เครื่องมือช่วยประมวลผลอัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID ซึ่งมี รูปแบบดังนี้ คือ Attribute1, Attribute2, Attribute3, ..., Attribute76 หมายถึงข้อมูลแต่ละ Attribute จะ เรียงต่อกันไป ดังแสดงในรูปที่ 3.1

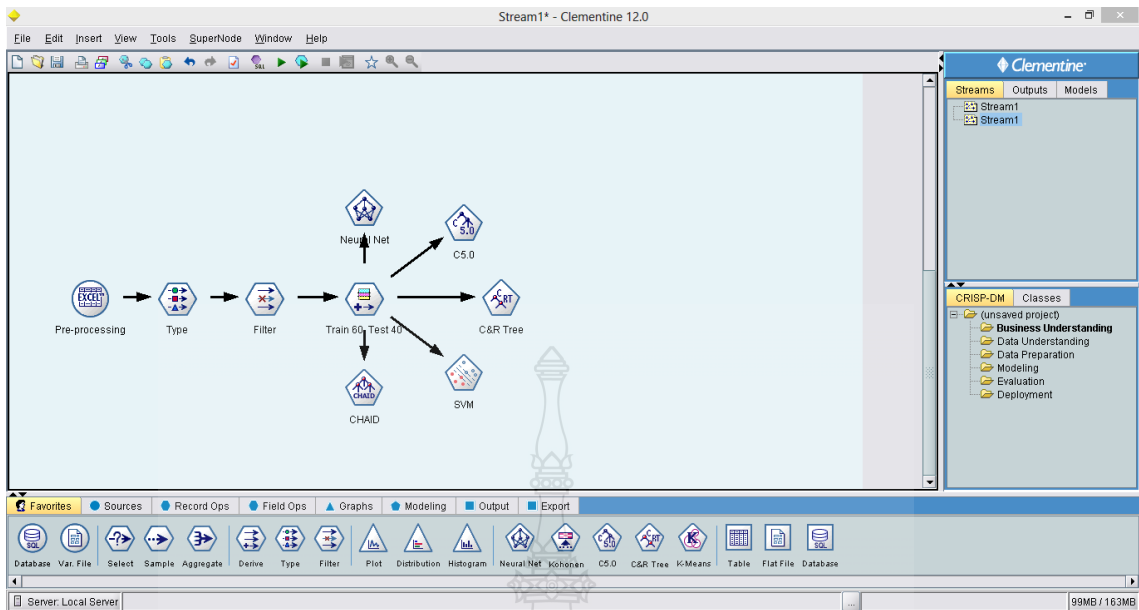
แอดทริบิวต์ต่างๆ ที่นำมาทดสอบ

จำนวนชุดข้อมูลที่น่ามาทดสอบ	No	Failure Mode	Attribute01	Attribute02	Attribute03	Attribute04	.....	Attribute76
	1	ATI/STE	data01	data02	data03	data04		
2	CND							
3	Degraded Head							
4	FHM							
5	Head Instability							
6	NMD							
7	NPF							
8	OTW							
9	VDW							
...	...							
1161	Weak write							

รูปที่ 3.1 โครงสร้างข้อมูลที่จะใช้กับ โปรแกรมประมวลผล

### 3.2.2 โปรแกรมที่ใช้ในการประมวลผล

เมื่อได้ข้อมูลมาตรฐานที่เหมาะสมแล้วจะนำข้อมูลมาตรฐานนี้ไปทำการทดสอบกับ อัลกอริทึมต่าง ๆ โดยในการทดสอบจะมีการปรับค่าสภาวะต่าง ๆ ของอัลกอริทึมให้เกิดความถูกต้อง เหมาะสมที่สุดจากการทำการประมวลผล โดยต้องอาศัยการประมวลผลข้อมูลจากอัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID ซึ่งเป็นอัลกอริทึมในการคัดแยก ปัจจุบันจำนวน โปรแกรมที่มีการพัฒนาอัลกอริทึมต่าง ๆ เหล่านี้ซึ่งใช้เป็นเครื่องมือช่วยในการคำนวณและวิเคราะห์ ความถูกต้องของข้อมูลมีมากขึ้นและมีการพัฒนาประสิทธิภาพของแต่ละอัลกอริทึมด้วย ซึ่งในการ ทดลองของการศึกษานี้ได้นำโปรแกรม Clementine 12.0 ซึ่งโปรแกรมที่มีประสิทธิภาพมาใช้เป็น เครื่องมือช่วยในการประมวลผลอัลกอริทึมต่าง ๆ ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 หน้าหลักของโปรแกรม Clementine 12.0

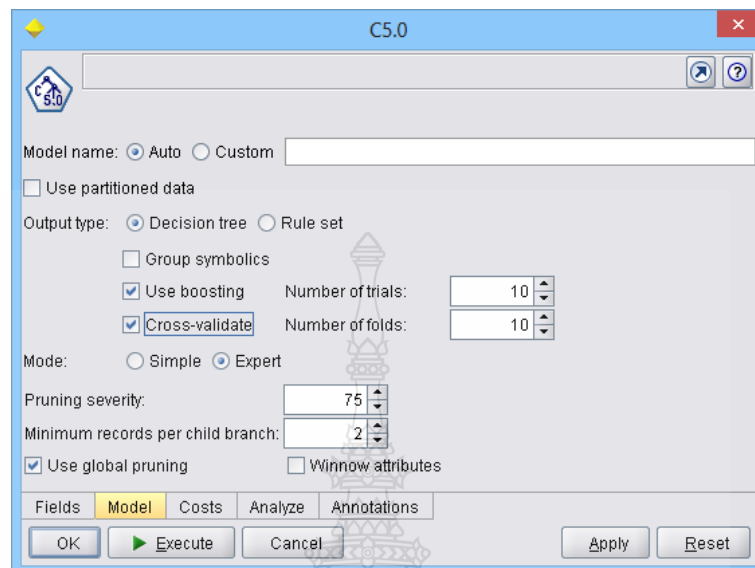
โปรแกรม Clementine 12.0 ถูกพัฒนาจากเวอร์ชันเดิมโดยประกอบด้วยรูปแบบและแบบจำลองที่เพิ่มมากขึ้น ทำให้ผู้ใช้สามารถนำไปใช้ทำงานได้อย่างมีประสิทธิภาพและได้ผลลัพธ์ที่มีความถูกต้องแม่นยำมากขึ้นด้วย ในการทดลองจะนำโปรแกรมนี้มาใช้ทดสอบอัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID ซึ่งผู้ใช้สามารถกำหนดค่าต่าง ๆ ของอัลกอริทึมได้เองทำให้ใช้งานง่ายขึ้น ใช้เวลาในการประมวลผล (time-to-solution) อย่างรวดเร็ว โปรแกรมสามารถทำการคำนวณผลลัพธ์ทั้งแบบไบนารี และแบบตัวเลขโดยใช้การดำเนินงานของ Automated modeling ซึ่งสามารถสร้างและประเมินผลแบบจำลองที่แตกต่างกันจำนวนมากได้ มีการใช้เทคนิคต่าง ๆ เพื่อให้การวิเคราะห์ของอัลกอริทึมแม่นยำยิ่งขึ้น เมื่อทำงานกับกลุ่มข้อมูลที่มีมากขึ้น และสามารถแสดงผลลัพธ์ได้หลายแบบ เช่น กราฟ ตาราง หรือ แผนภูมิ ทำให้วิเคราะห์และเข้าใจผลลัพธ์ได้ง่ายขึ้น

### 3.3 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม

เมื่อได้จัดเตรียมข้อมูลและมีโปรแกรมที่จะใช้ในการประมวลผลแล้วต้องปรับค่าต่าง ๆ ของโปรแกรม ทำได้โดยการนำโปรแกรมนั้นไปทดสอบกับข้อมูลที่นำมาสอนแล้วทำการปรับค่าของอัลกอริทึมให้มีสถานะที่เหมาะสมซึ่งทำให้ได้ผลลัพธ์ที่มีความถูกต้องแม่นยำมากที่สุดก่อนนำไปประมวลผลกับชุดข้อมูลที่นำมาทดสอบ

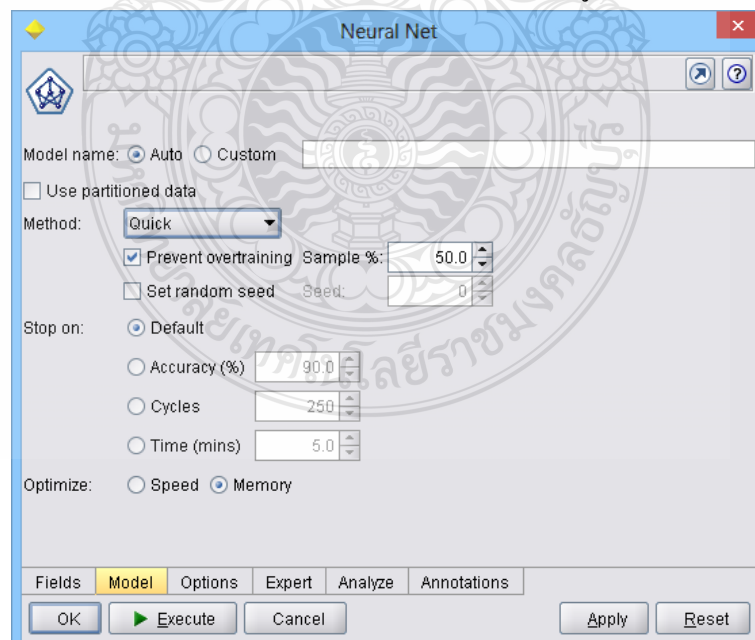
3.3.1 อัลกอริทึม C5.0 จะมีส่วนในการปรับค่าของอัลกอริทึมก่อนที่จะมีการสร้างแบบจำลอง โดยกำหนดให้มีการใช้งาน Boosting เพื่อเป็นการสร้างต้นไม้ตัดสินใจแบบคณะกรรมการ

และใช้การรวมผลลัพธ์ด้วยวิธี Average Voting ทั้งนี้จะมีการใช้งาน Cross-validation แบบ 10 fold ในการจำลองการสร้างแบบจำลองเสมือนการใช้งานจริง ดังแสดงในรูปที่ 3.3



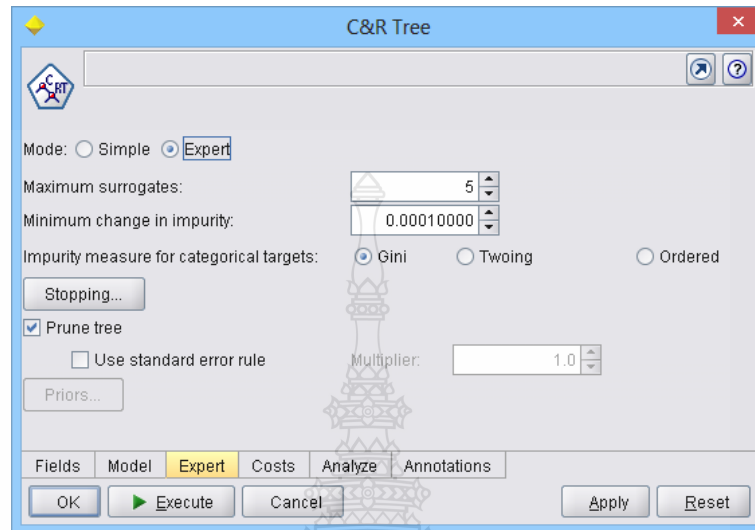
รูปที่ 3.3 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม C5.0

3.3.2 อัลกอริทึม Neural Network สามารถปรับเปลี่ยนค่าดังต่อไปนี้ กำหนดให้มีการสอนกระทั่งได้ค่าความถูกต้องที่ต้องการ กำหนดจำนวนรอบในการประมวลผล และกำหนดเวลาที่ใช้สำหรับการสอน ซึ่งจะทำการเลือกค่ามาตรฐานที่ใช้งาน ดังแสดงในรูปที่ 3.4



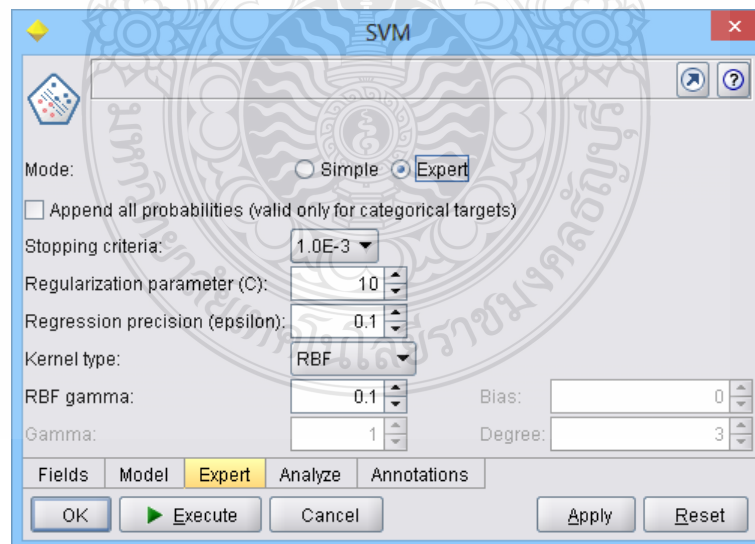
รูปที่ 3.4 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม Neural Network

3.3.3 อัลกอริทึม C&R Tree สามารถปรับเปลี่ยนค่า Maximum tree depth โดยเลือกให้มีค่าตั้งต้นเท่ากับ 5 และกำหนด Impurity measure เป็น Gini เพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องแม่นยำมากที่สุด ดังแสดงในรูปที่ 3.5



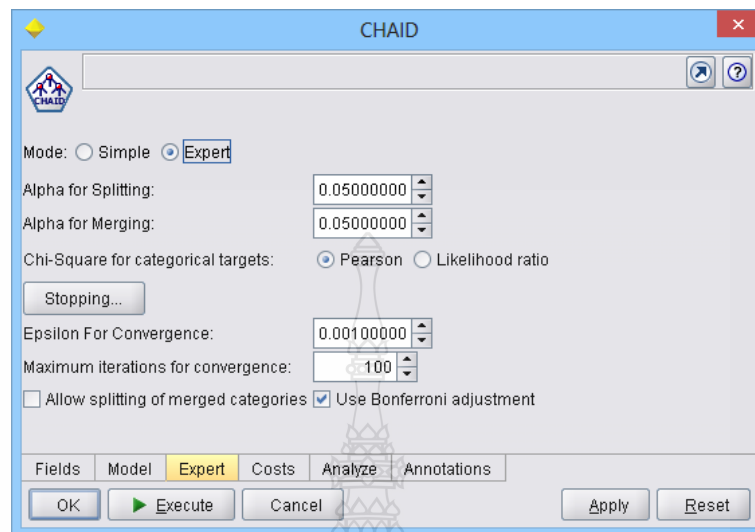
รูปที่ 3.5 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม C&R Tree

3.3.4 อัลกอริทึม SVM สามารถปรับเปลี่ยนค่า Stopping criteria สามารถกำหนดรูปแบบ Kernel type เป็น RBF, Polynomial, Sigmoid หรือ Linear โดยเลือกตามค่ามาตรฐานที่ใช้งาน ดังแสดงในรูปที่ 3.6



รูปที่ 3.6 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม SVM

3.3.5 อัลกอริทึม CHAID สามารถปรับเปลี่ยนค่าต่าง ๆ ได้แก่ Alpha for Splitting หรือ stopping criteria โดยเลือกตามค่าตั้งต้นที่ใช้งาน ดังแสดงในรูปที่ 3.7



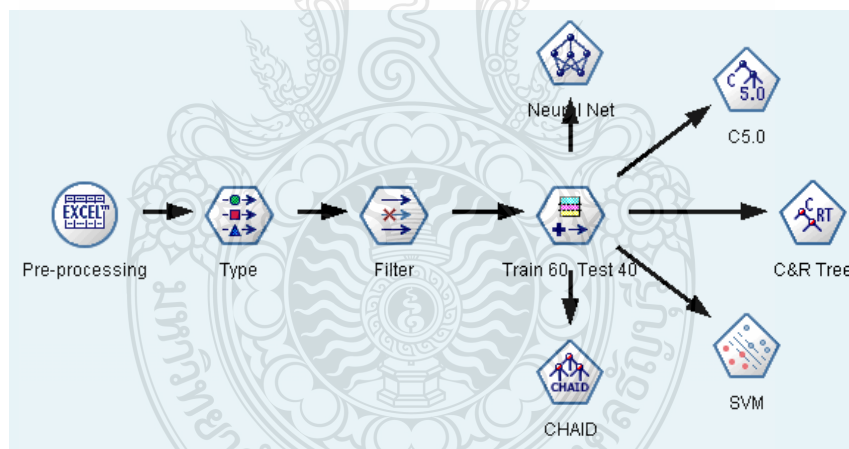
รูปที่ 3.7 การปรับแต่งค่าพารามิเตอร์ของอัลกอริทึม CHAID

### 3.4 ออกแบบการทดลอง

งานวิจัยนี้ออกแบบการทดลองสำหรับคัดแยกสาเหตุการเสียชีวิตทั้งหมดของฮาร์ดดิสก์จำนวน 1,161 ตัว ซึ่งมีสาเหตุการเสียชีวิตที่พบได้บ่อยในกระบวนการผลิตจำนวน 10 ประเภท โดยแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ ชุดข้อมูลสำหรับสอน (Training data) จำนวน 60% และชุดข้อมูลสำหรับทดสอบ (Testing data) จำนวน 40% ตามตารางที่ 3.4 แล้วใช้อัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID มาทำการทดสอบหาค่าความถูกต้องที่ได้จากแต่ละแบบจำลอง ซึ่งในการสร้างแบบจำลองนั้น จะนำข้อมูลที่ได้จากการเตรียมข้อมูลมาประมวลผล โดยเริ่มจากการกำหนดชนิดของแต่ละแอตทริบิวท์ที่ใช้ในการประมวลผลสำหรับอัลกอริทึมต่าง ๆ และทำการปรับแต่งค่าพารามิเตอร์ให้สอดคล้องกับการทดลองที่ได้ออกแบบไว้ จากนั้นจะทำการประมวลผลข้อมูลเพื่อสร้างแบบจำลองสำหรับคัดแยกสาเหตุการเสียชีวิตดังแสดงในรูปที่ 3.8

ตารางที่ 3.4 จำนวนชุดข้อมูลสำหรับสอน และชุดข้อมูลสำหรับทดสอบ

ประเภทอาการเสีย	ชุดข้อมูลสำหรับสอน	ชุดข้อมูลสำหรับทดสอบ	รวม
ATI/STE	5	4	9
CND	74	40	114
Degraded Head	18	10	28
FHM	48	19	67
Head Instability	30	16	46
NMD	184	146	330
NPF	257	182	439
OTW	10	6	16
VDW	45	45	90
Weak write	14	8	22
รวม	685	476	1,161



รูปที่ 3.8 กระบวนการสร้างแบบจำลองชนิดต่าง ๆ

วิธีการทดสอบความถูกต้อง และความน่าเชื่อถือของแบบจำลองนั้น ๆ จะต้องนำผลลัพธ์ที่แท้จริงของข้อมูลที่ใช้ทดสอบ (Testing data) มาเปรียบเทียบกับกลุ่มผลลัพธ์ที่หาได้จากแบบจำลอง การทำนาย โดยเลือกใช้ระบบการแบ่งแยกแบบ Confusion Matrix เพื่อแสดงผลและหาค่าความถูกต้องของผลที่ได้จากการทำนาย ดังแสดงในรูปที่ 3.9

		ค่าทำนาย (Predicted)									
		ATI/STE	CND	Degraded	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ค่าความจริง (Actual)	ATI/STE	4	0	0	0	0	0	0	0	0	0
	CND	0	40	0	0	0	0	0	0	0	0
	Degraded Head	0	0	9	0	0	0	0	1	0	0
	FHM	0	0	0	19	0	0	0	0	0	0
	Head Instability	0	0	0	0	16	0	0	0	0	0
	NMD	0	0	0	0	0	146	0	0	0	0
	NPF	0	0	0	0	0	0	182	0	0	0
	OTW	0	0	0	0	0	0	0	6	0	0
	VDW	0	0	0	0	0	0	1	1	43	0
	Weak write	0	0	0	0	0	0	0	0	0	8

รูปที่ 3.9 ตัวอย่างการหาค่าความถูกต้องของแบบจำลอง

ข้อมูลในแนวตั้ง คือ ผลลัพธ์ที่ได้จากการทำนาย จะขึ้นอยู่กับแบบจำลองที่ใช้ในการทำนาย สาเหตุการเสียนของฮาร์ดดิสก์ ว่ามีผลลัพธ์ที่แท้จริงอะไรบ้าง

ข้อมูลในแนวนอน คือ ผลลัพธ์ของข้อมูลที่แท้จริง จะขึ้นอยู่กับแบบจำลองที่ใช้ในการทำนาย สาเหตุการเสียนของฮาร์ดดิสก์ ว่ามีการกำหนดกลุ่มสาเหตุการเสียนของฮาร์ดดิสก์ไว้อะไรบ้าง

ซึ่งสามารถนำข้อมูลที่ได้ออกมาเพื่อหาค่าความถูกต้อง หรือค่าความเชื่อมั่นของผลการทำนายที่ได้จากแบบจำลองต่อไป

### 3.5 สรุปวิธีดำเนินงานวิจัย

ในขั้นตอนสุดท้ายหลังจากผ่านขั้นตอนการพัฒนาระบบผู้เชี่ยวชาญและทดสอบปรับปรุงจนระบบสามารถทำงานได้อย่างสมบูรณ์แล้ว การทดสอบการประมวลผลข้อมูลนี้ใช้การวัดความถูกต้องแม่นยำในการคัดแยกกลุ่มอาการเสียนของฮาร์ดดิสก์ แสดงค่าการวัดเป็นเปอร์เซ็นต์และพิจารณาผลโดยการเปรียบเทียบกันในแต่ละอัลกอริทึมด้วย การพิจารณาจะนำปัจจัยเรื่องเวลาพิจารณาไปด้วย โดยคิดเป็นการประมาณค่าเวลาในการประมวลผลของแต่ละอัลกอริทึม โดยพิจารณาระยะเวลาที่เริ่มมีการประมวลผลจนเสร็จสิ้นการประมวลผล เพื่อสำหรับนักวิเคราะห์ข้อมูลที่จะนำผลลัพธ์ที่ได้จากการทำนายมาสรุปความหมายกลายเป็นข้อมูลความรู้ซึ่งจะนำไปเป็นสารสนเทศที่ช่วยในการตัดสินใจต่อไป จากการวางแผนการดำเนินงานวิจัยเพื่อเป็นการเตรียมข้อมูล และอุปกรณ์ต่าง ๆ ที่ต้องใช้ในการทดลองช่วยทำให้รู้แนวทางในการทดลอง การวิเคราะห์ผลและวัดประสิทธิภาพซึ่งเป็นการเตรียมความพร้อมก่อนเข้าสู่การทำการทดลองที่มีขั้นตอนต่าง ๆ ในบทต่อไป



## บทที่ 4

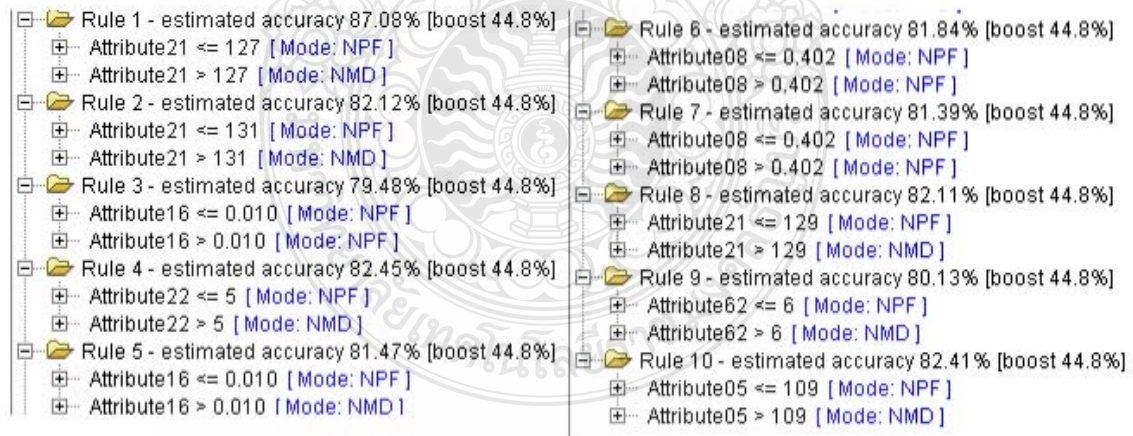
### ผลการทดลอง

ในบทนี้จะนำเสนอผลการดำเนินงานในแต่ละขั้นตอนซึ่งประกอบด้วย ตัวแบบในการทำนายอาการเสียของฮาร์ดดิสก์ ผลการทดสอบตัวแบบ กฎการจำแนก ซึ่งได้ผ่านการประมวลผลแต่ละอัลกอริทึมจนครบและทำการบันทึกผลในแต่ละครั้งที่ทำการทดสอบ วิเคราะห์และเปรียบเทียบผลลัพธ์ที่ได้จากการทดสอบในแต่ละอัลกอริทึม โดยมีผลการทดลองซึ่งนำเสนอค่าความถูกต้องในรูปแบบเปอร์เซ็นต์ (%) แล้วทำการสรุปผลการทดลอง

#### 4.1 ผลการสร้างแบบจำลอง

เพื่อทดลองหาคุณลักษณะของแต่ละแบบจำลองว่าใช้เวลาเท่าใดในการสร้างแบบจำลอง ให้มีความสำคัญกับแอตทริบิวต์อะไรบ้างในการวิเคราะห์ จึงได้ทดลองปรับแต่งค่าในรูปแบบต่าง ๆ แล้วมาทำการฝึกสอนและทดสอบ ซึ่งได้ผลลัพธ์ดังนี้

4.1.1 อัลกอริทึม C5.0 มีการสร้างความลึก (Tree depth) ทั้งหมดจำนวน 45 ชั้น กำหนดให้แอตทริบิวต์ Failure Mode เป็นแอตทริบิวต์เป้าหมาย (Target) และใช้เวลาในการสร้างแบบจำลอง 18 วินาที โดยมีกำหนดความสำคัญของแอตทริบิวต์ที่นำมาใช้ประมวลผล (Input) ดังแสดงในรูปที่ 4.1

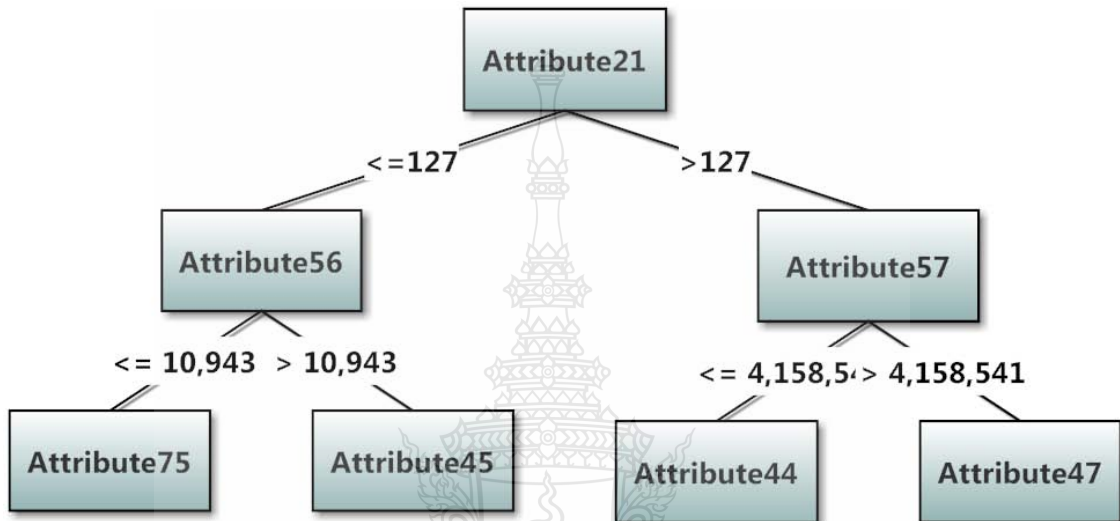


รูปที่ 4.1 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง C5.0

จากรูปที่ 4.1 แบบจำลองที่ได้จากอัลกอริทึม C5.0 สามารถแบ่งเป็น โมเดล (Rule) ออกเป็น 10 โมเดล โดยที่โมเดลแรกจะเริ่มต้นประมวลผลข้อมูลเพื่อการทำนาย หากโมเดลแรกมีส่วนที่ทำนายผิดพลาด โมเดลที่สองจะนำข้อมูลส่วนที่โมเดลแรกทำนายผิดพลาดมาประมวลผลใหม่

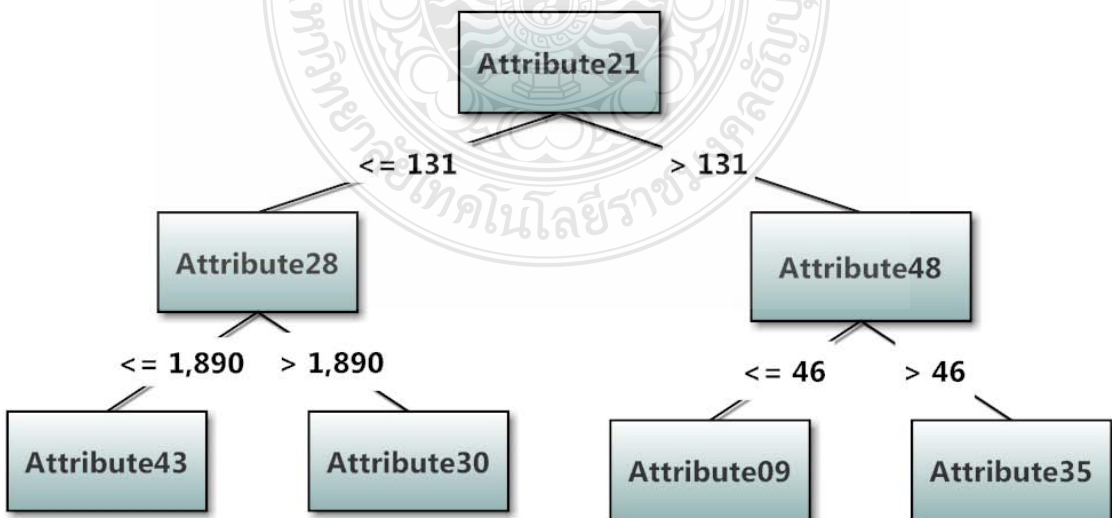
โมเดลที่สามจะนำข้อมูลส่วนที่โมเดลที่สองทำนายผิดพลาดมาประมวลผลใหม่ ซึ่งทำเช่นนี้ต่อไปจนครบทั้ง 10 โมเดล ดังนี้

4.1.1.1 โมเดลที่ 1 มีความลึก 35 ชั้น โดยมีการเรียกใช้ Attribute21 เป็นโหนดราก ซึ่งมี Attribute56 และ Attribute57 ในลำดับชั้นที่ 2 และมี Attribute75, Attribute45, Attribute44 และ Attribute47 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.2



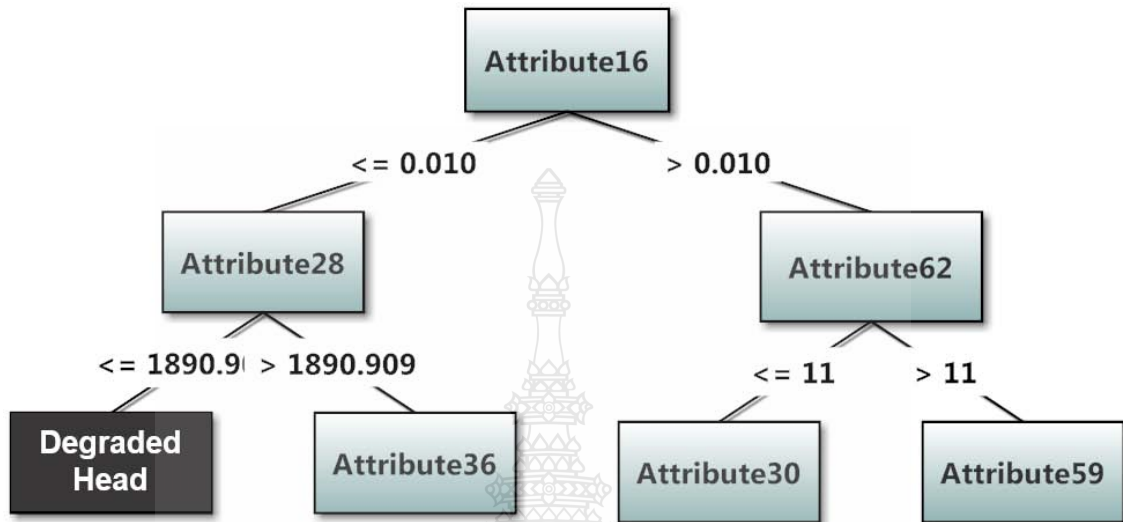
รูปที่ 4.2 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 1

4.1.1.2 โมเดลที่ 2 มีความลึก 42 ชั้น โดยมีการเรียกใช้ Attribute21 เป็นโหนดราก ซึ่งมี Attribute28 และ Attribute48 ในลำดับชั้นที่ 2 และมี Attribute43, Attribute30, Attribute09 และ Attribute35 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.3



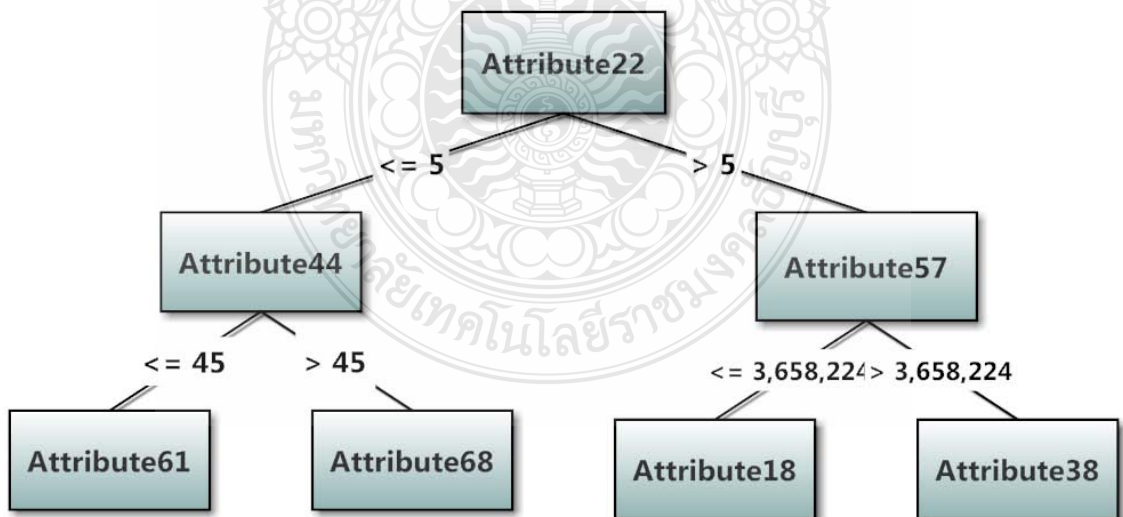
รูปที่ 4.3 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 2

4.1.1.3 โมเดลที่ 3 มีความลึก 32 ชั้น โดยมีการเรียกใช้ Attribute16 เป็นโหนดราก ซึ่งมี Attribute28 และ Attribute62 ในลำดับชั้นที่ 2 และมี Attribute36, Attribute30 และ Attribute59 ในลำดับชั้นที่ 3 และสามารถแยกสาเหตุการเสียได้ คือ Degraded Head ดังแสดงในรูปที่ 4.4



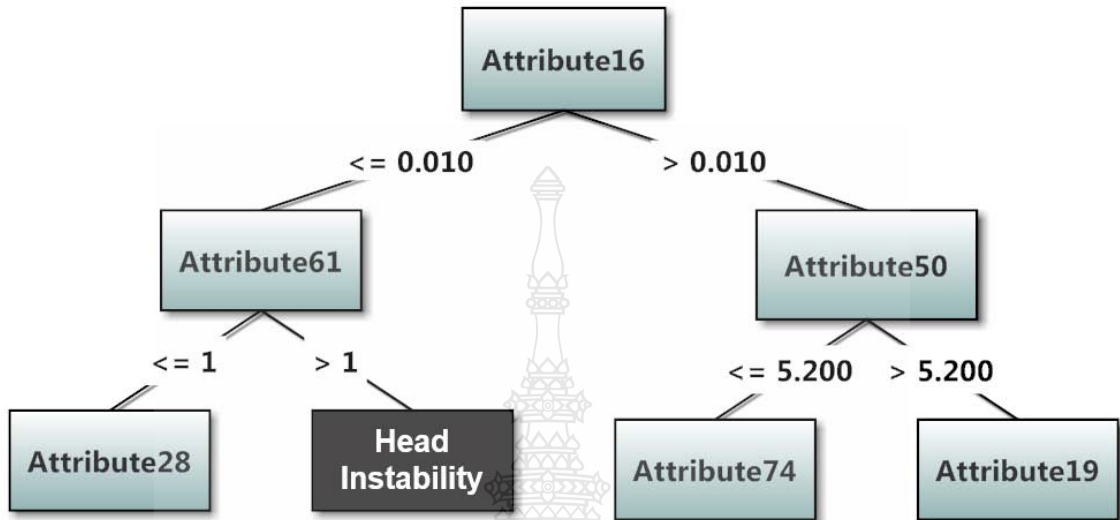
รูปที่ 4.4 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 3

4.1.1.4 โมเดลที่ 4 มีความลึก 35 ชั้น โดยมีการเรียกใช้ Attribute22 เป็นโหนดราก ซึ่งมี Attribute44 และ Attribute57 ในลำดับชั้นที่ 2 และมี Attribute61, Attribute68, Attribute18 และ Attribute38 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.5



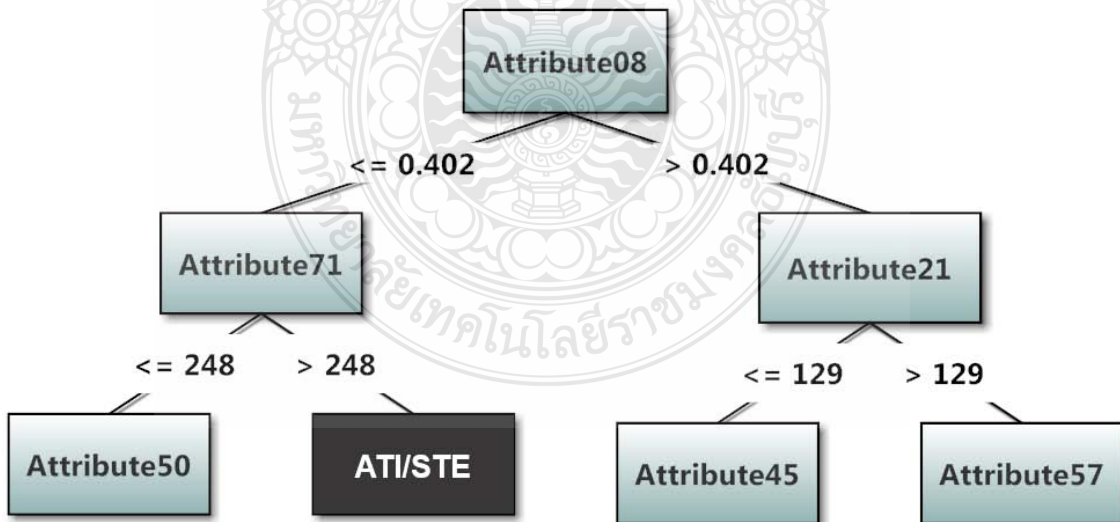
รูปที่ 4.5 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 4

4.1.1.5 โมเดลที่ 5 มีความลึก 34 ชั้น โดยมีการเรียกใช้ Attribute16 เป็นโหนดราก ซึ่งมี Attribute61 และ Attribute50 ในลำดับชั้นที่ 2 และมี Attribute28, Attribute74 และ Attribute19 ในลำดับชั้นที่ 3 และสามารถแยกสาเหตุการเสียได้ คือ Head Instability ดังแสดงในรูปที่ 4.6



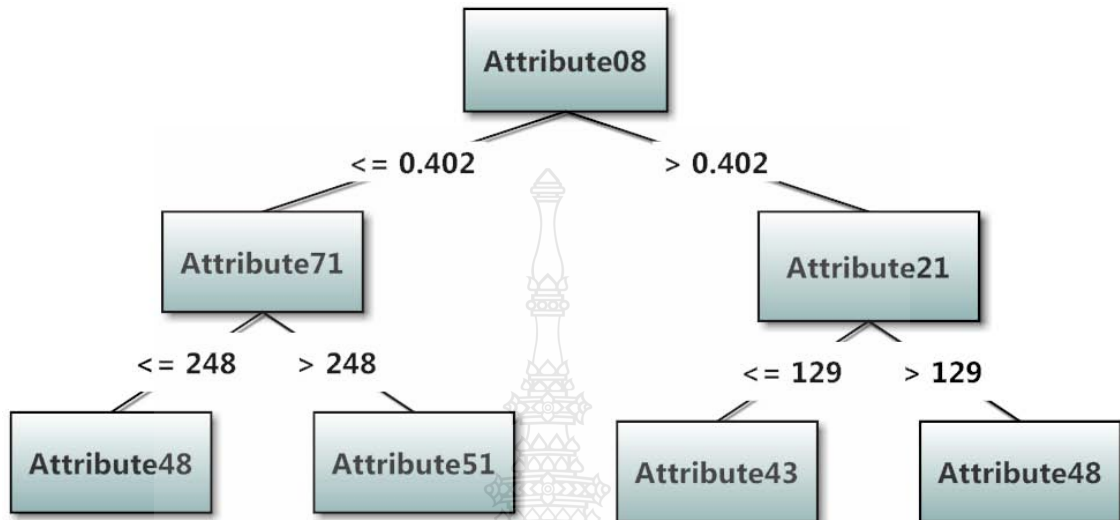
รูปที่ 4.6 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 5

4.1.1.6 โมเดลที่ 6 มีความลึก 39 ชั้น โดยมีการเรียกใช้ Attribute08 เป็นโหนดราก ซึ่งมี Attribute71 และ Attribute21 ในลำดับชั้นที่ 2 และมี Attribute50, Attribute45 และ Attribute57 ในลำดับชั้นที่ 3 และสามารถแยกสาเหตุการเสียได้ คือ ATI/STE ดังแสดงในรูปที่ 4.7



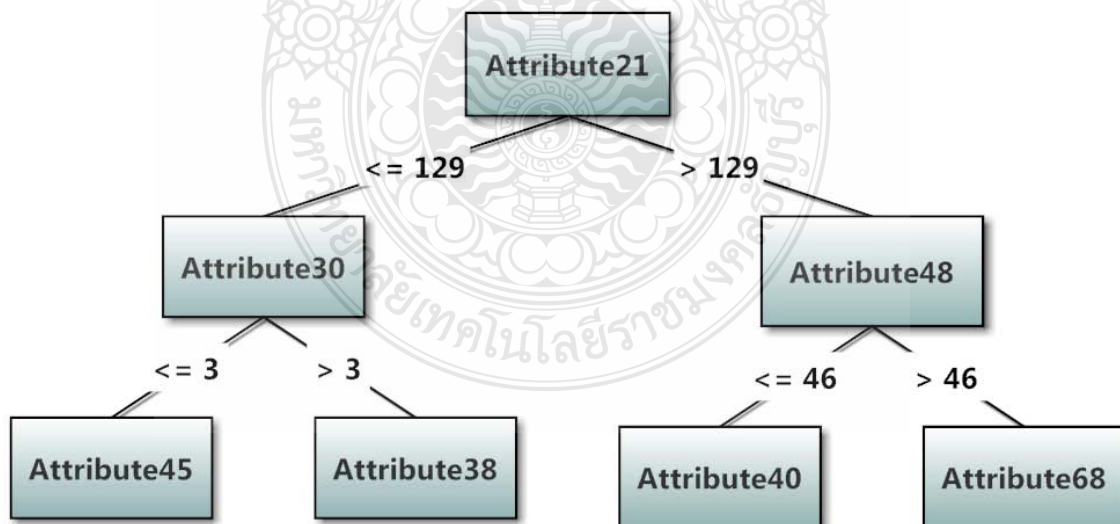
รูปที่ 4.7 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 6

4.1.1.7 โมเดลที่ 7 มีความลึก 39 ชั้น โดยมีการเรียกใช้ Attribute08 เป็นโหนดราก ซึ่งมี Attribute71 และ Attribute21 ในลำดับชั้นที่ 2 และมี Attribute48, Attribute51, Attribute43 และ Attribute48 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.8



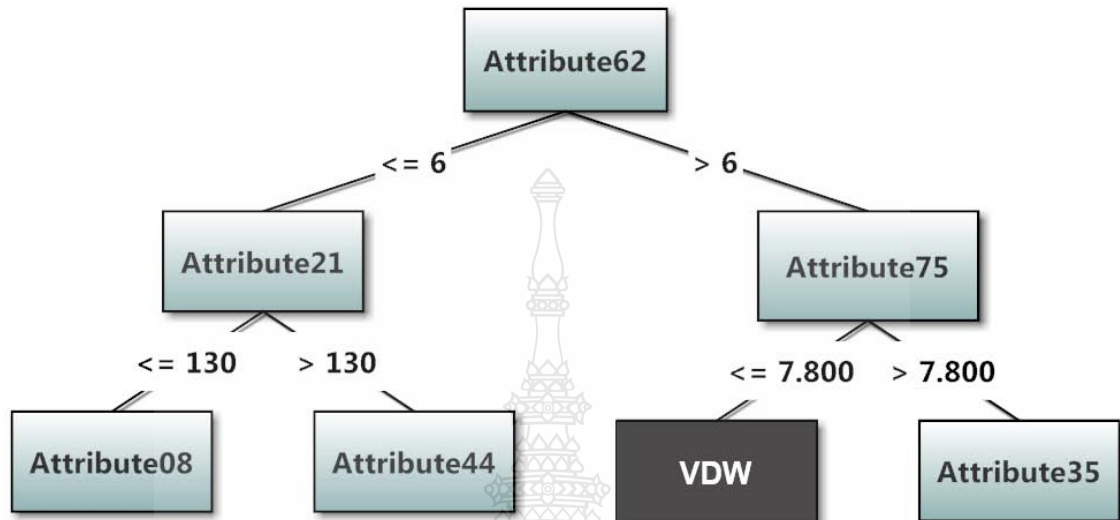
รูปที่ 4.8 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 7

4.1.1.8 โมเดลที่ 8 มีความลึก 44 ชั้น โดยมีการเรียกใช้ Attribute21 เป็นโหนดราก ซึ่งมี Attribute30 และ Attribute48 ในลำดับชั้นที่ 2 และมี Attribute45, Attribute38, Attribute40 และ Attribute68 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.9



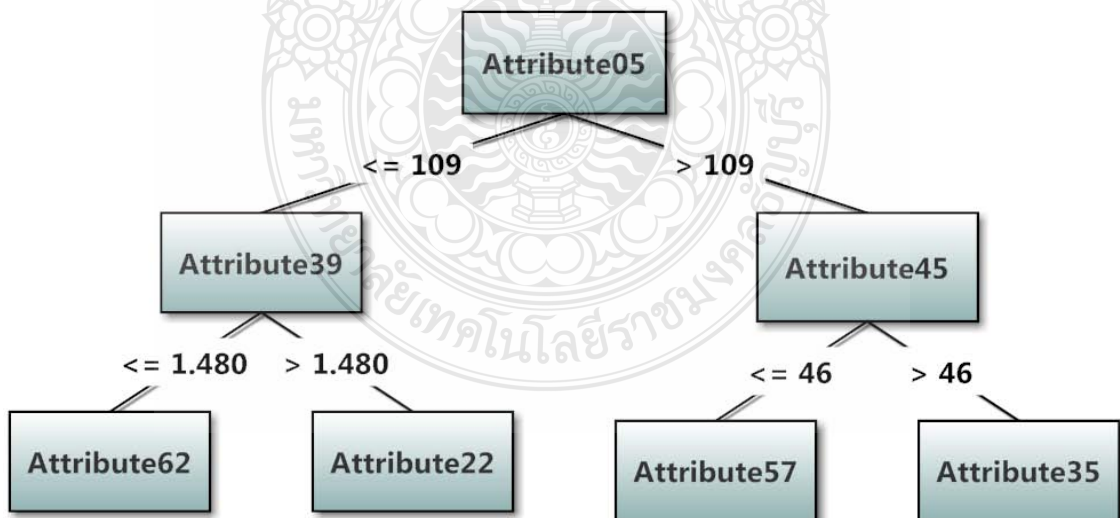
รูปที่ 4.9 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 8

4.1.1.9 โมเดลที่ 9 มีความลึก 35 ชั้น โดยมีการเรียกใช้ Attribute62 เป็น โหนดราก ซึ่งมี Attribute21 และ Attribute75 ในลำดับชั้นที่ 2 และมี Attribute08, Attribute44 และ Attribute35 ในลำดับชั้นที่ 3 และสามารถแยกสาเหตุการเสียได้ คือ VDW ดังแสดงในรูปที่ 4.10



รูปที่ 4.10 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 9

4.1.1.10 โมเดลที่ 10 มีความลึก 39 ชั้น โดยมีการเรียกใช้ Attribute05 เป็น โหนดราก ซึ่งมี Attribute39 และ Attribute45 ในลำดับชั้นที่ 2 และมี Attribute62, Attribute22, Attribute57 และ Attribute35 ในลำดับชั้นที่ 3 ดังแสดงในรูปที่ 4.11



รูปที่ 4.11 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของโมเดลที่ 10

จากกฎของแบบจำลอง C5.0 ทั้ง 10 โมเดลพบว่า แอตทริบิวต์ที่มีความสำคัญในลำดับที่ 1 ที่ถูกกำหนดโดยผู้เชี่ยวชาญ มีการเรียกใช้งานจากแบบจำลองใน 5 ลำดับชั้นแรกของความสำคัญ ดังตารางที่ 4.1 ซึ่งทำให้ค่าความถูกต้องที่ได้มีความน่าเชื่อถือเนื่องจากมีความสอดคล้องกับการทำงานจริงของผู้เชี่ยวชาญ และพบว่าแอตทริบิวต์บางแอตทริบิวต์ไม่ถูกนำมาใช้งาน ได้แก่ Attribute29, Attribute46 และ Attribute49 รวมถึงแอตทริบิวต์ที่มีการเรียกใช้งานเพียง 1-3 โมเดลเท่านั้น และไม่ได้อยู่ในลำดับความสำคัญ 1-5 ของแต่ละโมเดล ได้แก่ Attribute42 และ Attribute52 ซึ่งแอตทริบิวต์เหล่านี้อาจสามารถตัดออกจากแบบจำลองเพื่อเพิ่มความเร็วในการประมวลผลและลดการใช้งานหน่วยความจำลง เนื่องจากเป็นแอตทริบิวต์ที่มีความสำคัญน้อยสำหรับการคัดแยกสาเหตุการเสียชีวิตของฮาร์ดดิสก์

ตารางที่ 4.1 แสดงแอตทริบิวต์ที่มีการใช้งานในกระบวนการสร้างแบบจำลอง C5.0

Attribute	Priority	Rule									
		1	2	3	4	5	6	7	8	9	10
Attribute08	1	5	●	●	●	●	1	1	●	3	●
Attribute09	1	●	3	5	●	●	●	●	●	●	●
Attribute16	1	●	5	1	●	1	●	●	●	●	●
Attribute17	1	5	●	●	●	●	●	●	●	●	●
Attribute18	1	●	●	●	3	●	●	●	●	●	●
Attribute19	1	●	●	●	●	3	●	●	●	●	●
Attribute20	1	●	●	●	●	●	●	4	5	X	●
Attribute21	1	1	1	●	4	●	2	2	1	2	●
Attribute22	1	4	4	●	1	●	4	●	●	●	3
Attribute32	1	●	●	4	4	●	●	●	●	●	X
Attribute35	1	●	3	●	●	5	●	●	●	3	3
Attribute41	1	●	●	●	●	●	●	●	●	●	5
Attribute43	1	●	3	●	●	●	●	3	●	4	●
Attribute44	1	3	●	●	2	●	5	●	●	3	●
Attribute45	1	3	X	●	●	●	3	●	3	●	2

ตารางที่ 4.1 แสดงแอตทริบิวต์ที่มีการใช้งานในกระบวนการสร้างแบบจำลอง C5.0 (ต่อ)

Attribute	Priority	Rule									
		1	2	3	4	5	6	7	8	9	10
Attribute04	2	4	●	●	●	●	●	●	●	5	●
Attribute05	2	●	●	●	●	●	●	●	5	●	1
Attribute12	2	●	●	●	●	●	●	●	4	5	●
Attribute13	2	●	5	●	5	●	5	●	●	4	5
Attribute27	2	●	●	●	●	●	●	●	●	●	●
Attribute28	2	●	2	2	●	3	●	●	X	●	●
Attribute30	2	●	3	3	●	4	●	●	2	●	●
Attribute33	2	4	●	●	●	●	4	5	●	●	4
Attribute34	2	●	●	●	●	●	●	●	●	X	●
Attribute38	2	●	4	●	3	X	5	5	3	●	5
Attribute39	2	●	●	5	●	X	●	●	●	●	2
Attribute40	2	●	4	4	●	●	●	4	3	●	●
Attribute74	2	●	5	5	4	3	4	5	4	●	●
Attribute75	2	3	●	●	●	●	●	X	●	2	●
Attribute76	2	●	●	●	●	●	●	●	●	●	●
Attribute06	3	X	●	●	●	●	4	●	4	●	●
Attribute07	3	●	●	●	X	●	5	●	●	●	●
Attribute14	3	X	●	●	X	●	●	●	●	4	●
Attribute23	3	4	●	●	●	●	●	5	5	●	●
Attribute24	3	X	●	●	●	●	●	●	●	●	●
Attribute25	3	●	●	●	●	●	●	X	●	●	●
Attribute26	3	●	●	5	●	●	5	5	4	●	●
Attribute36	3	5	●	3	●	4	●	X	●	X	●
Attribute42	3	●	X	●	X	X	X	●	X	X	X
Attribute55	3	●	●	●	●	5	●	●	●	●	4



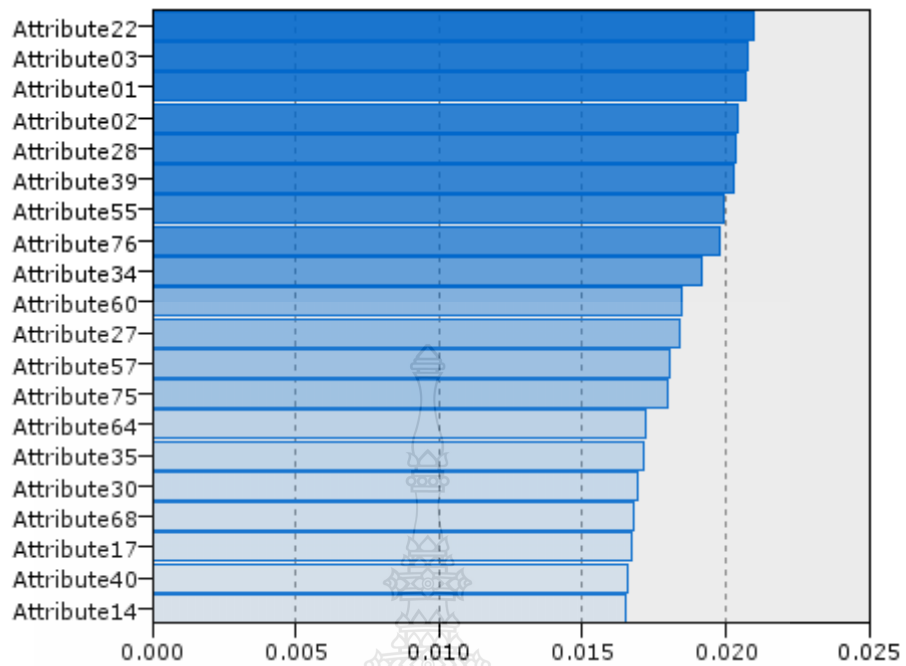
ตารางที่ 4.1 แสดงแอตทริบิวต์ที่มีการใช้งานในกระบวนการสร้างแบบจำลอง C5.0 (ต่อ)

Attribute	Priority	Rule									
		1	2	3	4	5	6	7	8	9	10
Attribute56	3	2	●	X	4	5	●	●	●	●	●
Attribute57	3	2	X	5	2	●	3	4	●	●	3
Attribute58	3	4	●	●	●	●	X	●	●	●	●
Attribute59	3	●	●	3	●	X	●	●	X	X	●
Attribute60	3	●	X	5	5	4	●	●	5	●	●
Attribute61	3	●	●	X	3	2	X	●	●	●	4
Attribute62	3	●	5	2	●	●	●	●	●	1	3
Attribute01	4	●	●	●	●	●	●	●	●	●	●
Attribute02	4	●	4	●	5	●	●	●	5	X	●
Attribute03	4	5	●	●	●	4	●	●	●	5	●
Attribute63	4	●	5	●	●	●	●	●	4	5	●
Attribute64	4	●	●	●	●	●	5	●	●	●	●
Attribute65	4	●	●	●	●	●	●	●	●	●	●
Attribute66	4	●	●	●	●	●	●	●	●	●	X
Attribute67	4	●	●	●	●	●	●	●	●	●	●
Attribute68	4	●	5	●	3	●	●	4	3	●	●
Attribute69	4	●	●	●	●	●	●	●	●	●	5
Attribute70	4	●	●	●	5	●	●	●	●	●	●
Attribute10	5	●	●	●	●	●	●	●	●	X	●
Attribute11	5	●	●	●	●	●	●	●	●	●	●
Attribute15	5	●	●	●	●	●	X	●	●	●	●
Attribute29	5	X	X	X	X	X	X	X	X	X	X
Attribute31	5	●	●	●	●	●	●	●	●	●	●
Attribute37	5	●	●	●	●	●	●	●	●	X	●
Attribute46	5	X	X	X	X	X	X	X	X	X	X

ตารางที่ 4.1 แสดงแอตทริบิวต์ที่มีการใช้งานในกระบวนการสร้างแบบจำลอง C5.0 (ต่อ)

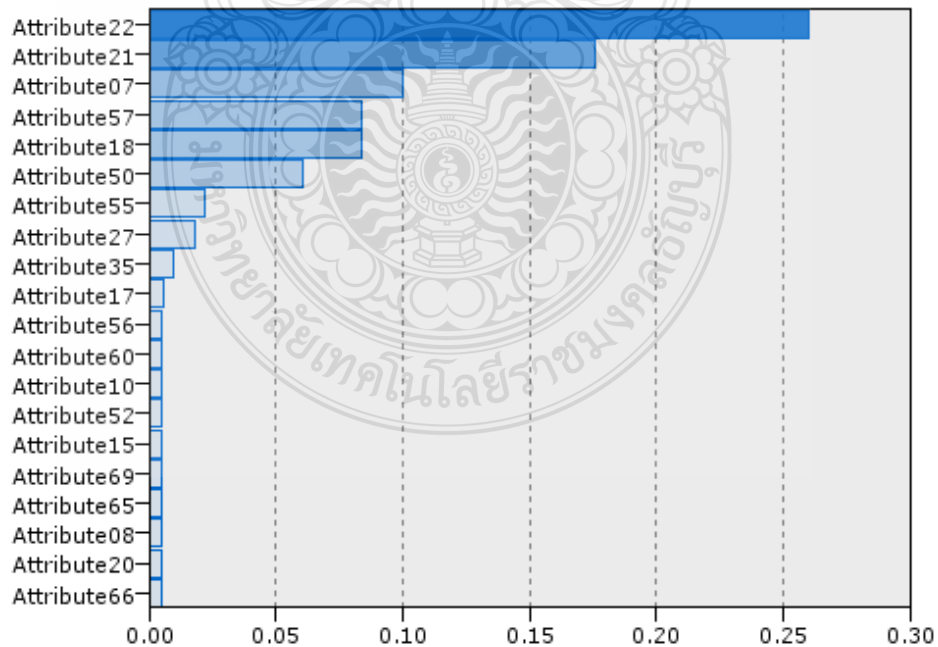
Attribute	Priority	Rule									
		1	2	3	4	5	6	7	8	9	10
Attribute47	5	3	5	●	X	●	4	●	●	●	●
Attribute48	5	●	2	●	●	●	●	3	2	5	●
Attribute49	5	X	X	X	X	X	X	X	X	X	X
Attribute50	5	●	●	4	●	2	3	●	●	●	●
Attribute51	5	●	●	●	4	●	X	3	X	●	●
Attribute52	5	X	X	X	X	X	●	X	X	X	X
Attribute53	5	●	●	●	●	●	●	●	●	●	●
Attribute54	5	●	4	●	●	●	●	●	●	●	●
Attribute71	5	●	●	X	●	X	2	2	●	●	5
Attribute72	5	●	X	X	X	X	X	X	X	X	4
Attribute73	5	●	●	●	●	5	●	5	4	4	●

4.1.2 อัลกอริทึม Neural Network มีการสร้างชั้นซ่อน (Hidden Layer) จำนวน 40 นิวรอน มีค่าประมาณการความถูกต้องระหว่างทำการฝึกสอนเท่ากับ 90.784% ใช้เวลาในการสร้างแบบจำลอง 1 นาที 38 วินาที โดยมีการกำหนดความสำคัญของแอตทริบิวต์ที่นำมาใช้ประมวลผล (Input) ดังแสดงในรูปที่ 4.12



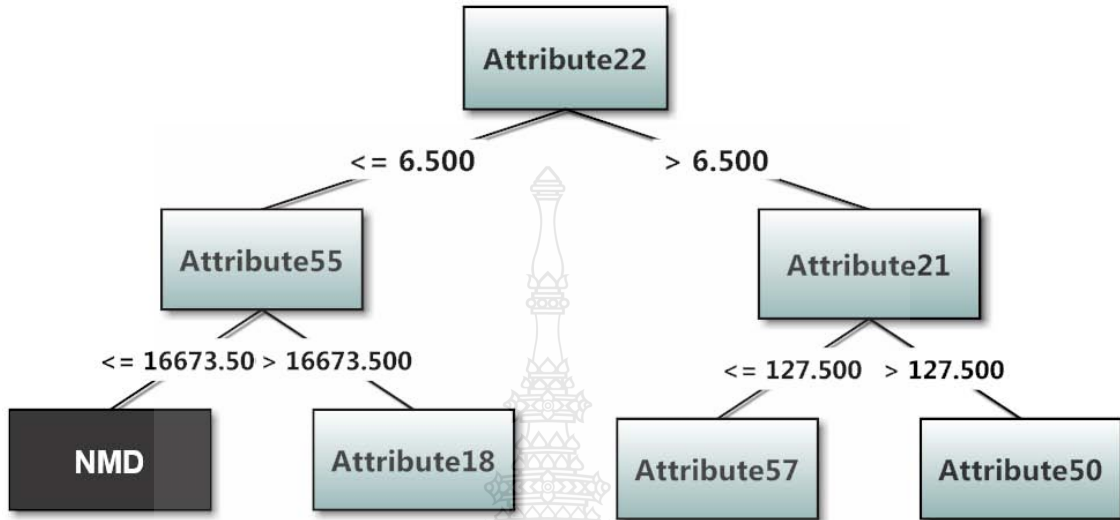
รูปที่ 4.12 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง Neural Network

4.1.3 อัลกอริทึม C&R Tree มีการสร้างความลึก (Tree depth) ทั้งหมดจำนวน 7 ชั้น กำหนดให้แอตทริบิวต์ Failure Mode เป็นแอตทริบิวต์เป้าหมาย และใช้เวลาในการสร้างแบบจำลอง 6 วินาที โดยมีกำหนดความสำคัญของแอตทริบิวต์ที่นำมาใช้ประมวลผล (Input) ดังแสดงในรูปที่ 4.13



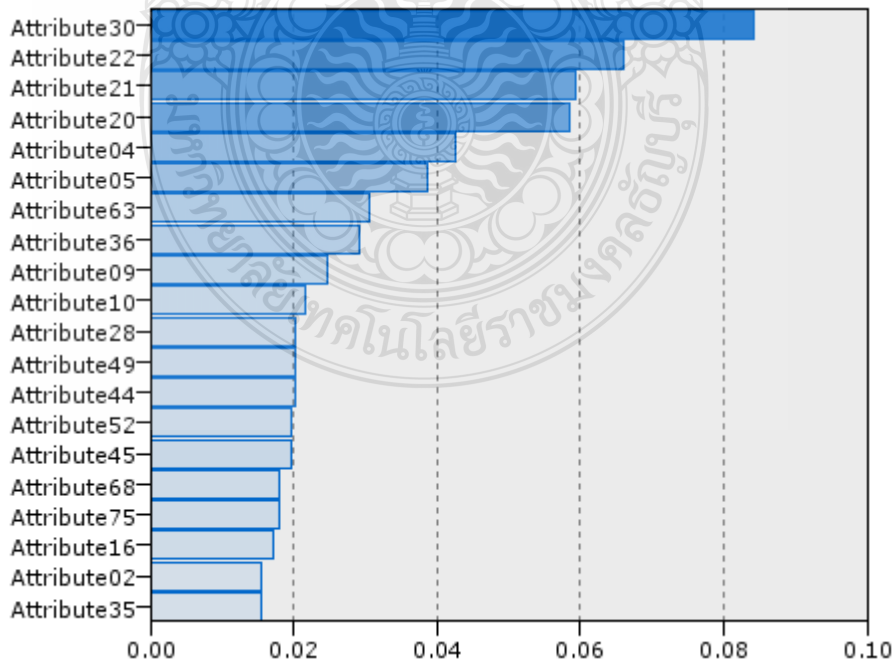
รูปที่ 4.13 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง C&R Tree

แบบจำลองที่ได้จากอัลกอริทึม C&R Tree มีการเรียกใช้ Attribute22 เป็นโหนดราก ซึ่งมี Attribute55 และ Attribute21 ในลำดับชั้นที่ 2 และมี Attribute18, Attribute57 และ Attribute50 ในลำดับชั้นที่ 3 และสามารถแยกสาเหตุการเสียได้ คือ VDW ดังแสดงในรูปที่ 4.14



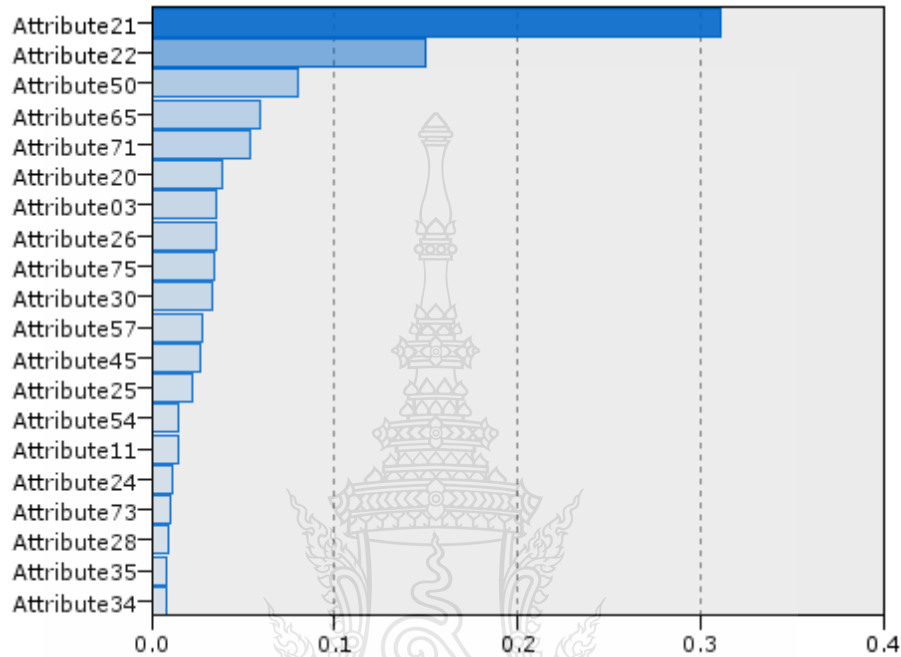
รูปที่ 4.14 แอตทริบิวท์ที่ใช้ใน 3 ลำดับชั้นแรกของแบบจำลอง C&R Tree

4.1.4 อัลกอริทึม SVM มีการกำหนดให้แอตทริบิวท์ Failure Mode เป็นแอตทริบิวท์เป้าหมาย (Target) และใช้เวลาในการสร้างแบบจำลอง 13 วินาที โดยแบบจำลองได้มีกำหนดความสำคัญของแอตทริบิวท์ที่นำมาใช้ประมวลผล (Input) ดังแสดงในรูปที่ 4.15



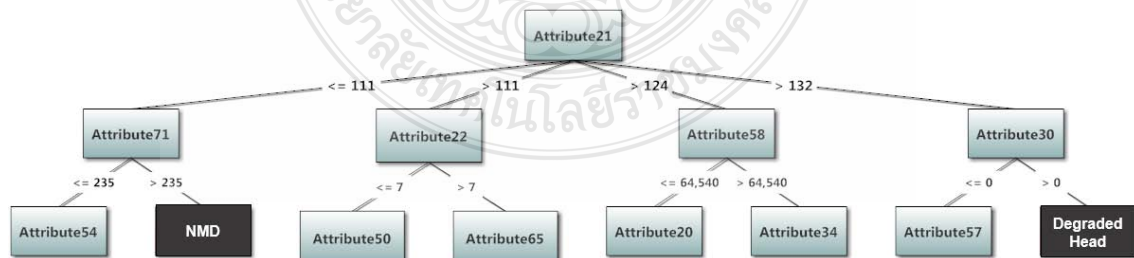
รูปที่ 4.15 ลำดับความสำคัญของแอตทริบิวท์ที่นำมาประมวลผลแบบจำลอง SVM

4.1.5 อัลกอริทึม CHAID มีการสร้างความลึก (Tree depth) ทั้งหมดจำนวน 8 ชั้น กำหนดให้แอตทริบิวต์ Failure Mode เป็นแอตทริบิวต์เป้าหมาย (Target) และใช้เวลาในการสร้างแบบจำลอง 3 วินาที โดยมีกำหนดความสำคัญของแอตทริบิวต์ที่นำมาใช้ประมวลผล (Input) ดังแสดงในรูปที่ 4.16



รูปที่ 4.16 ลำดับความสำคัญของแอตทริบิวต์ที่นำมาประมวลผลแบบจำลอง CHAID

แบบจำลองที่ได้จากอัลกอริทึม CHAID ได้มีการเรียกใช้และจัดความสำคัญ Attribute21 เป็นโหนดราก ซึ่งมี Attribute71, Attribute22, Attribute58 และ Attribute30 ในลำดับชั้นที่ 2 และมีการใช้ Attribute54, Attribute50, Attribute65, Attribute20, Attribute34 และ Attribute57 ในลำดับชั้นที่ 3 โดยสามารถแยกสาเหตุการเสียได้ คือ NPF และ Degraded Head ดังแสดงในรูปที่ 4.17



รูปที่ 4.17 แอตทริบิวต์ที่ใช้ใน 3 ลำดับชั้นแรกของแบบจำลอง CHAID

## 4.2 การทดลองหาประสิทธิภาพของแบบจำลอง

เป็นการวิเคราะห์ข้อมูลหลังจากได้รูปแบบของแบบจำลองที่ต้องการแล้วเพื่อหาค่าความถูกต้องของแต่ละแบบจำลอง โดยวิเคราะห์หาค่าความถูกต้องของชุดข้อมูลที่นำมาใช้สำหรับสอน (Training data) และหาค่าความถูกต้องของชุดข้อมูลที่นำมาใช้สำหรับทดสอบ (Testing data) ซึ่งได้ผลลัพธ์ดังนี้

4.2.1 อัลกอริทึม C5.0 ผลการวิเคราะห์ข้อมูลการทำนายผลการเรียนด้วยข้อมูลสำหรับสอน (Training data) จำนวน 685 เรคคอร์ด มีความสามารถในการเรียนรู้ได้ถูกต้อง เป็นจำนวน 677 เรคคอร์ด คิดเป็น 98.83% และผลการเรียนรู้ไม่ถูกต้อง จำนวน 8 เรคคอร์ด คิดเป็น 1.17% และผลการทำนายด้วยข้อมูลทดสอบ (Testing Data) จำนวน 476 เรคคอร์ด มีความสามารถในการทำนายได้ถูกต้อง เป็นจำนวน 475 เรคคอร์ด คิดเป็น 99.79% และผลการทำนายไม่ถูกต้อง จำนวน 1 เรคคอร์ด คิดเป็น 0.21% จากข้อมูลทั้งหมด สามารถอธิบายผลที่เป็นรูปแบบ Confusion Matrix ดังแสดงในรูปที่ 4.18 และแสดงค่าความถูกต้องของแต่ละสาเหตุการเสียของแบบจำลองได้ตามตารางที่ 4.2

Results for output field Failure Mode

Comparing \$C-Failure Mode with Failure Mode

'Partition'	1_Training	2_Testing
Correct	677 98.83%	475 99.79%
Wrong	8 1.17%	1 0.21%
Total	685	476

Coincidence Matrix for \$C-Failure Mode (rows show actuals)

'Partition' = 1_Training	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	5	0	0	0	0	0	0	0	0	0
CND	0	74	0	0	0	0	0	0	0	0
Degraded Head	0	0	16	0	0	0	2	0	0	0
FHM	0	0	0	47	0	1	0	0	0	0
Head Instability	0	0	0	0	30	0	0	0	0	0
NMD	0	0	0	0	0	182	2	0	0	0
NPF	0	0	0	0	0	0	257	0	0	0
OTW	0	0	0	0	0	0	0	10	0	0
VDW	0	0	0	0	0	0	1	0	44	0
Weak write	0	0	0	0	0	0	2	0	0	12
'Partition' = 2_Testing	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	4	0	0	0	0	0	0	0	0	0
CND	0	40	0	0	0	0	0	0	0	0
Degraded Head	0	0	9	0	0	0	1	0	0	0
FHM	0	0	0	19	0	0	0	0	0	0
Head Instability	0	0	0	0	16	0	0	0	0	0
NMD	0	0	0	0	0	146	0	0	0	0
NPF	0	0	0	0	0	0	182	0	0	0
OTW	0	0	0	0	0	0	0	6	0	0
VDW	0	0	0	0	0	0	0	0	45	0
Weak write	0	0	0	0	0	0	0	0	0	8

รูปที่ 4.18 ผลการทดสอบของแบบจำลอง C5.0

ตารางที่ 4.2 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของแบบจำลอง C5.0

Failure Mode	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
ATI/STE	5	0	100.00%	4	0	100.00%
CND	74	0	100.00%	40	0	100.00%
Degraded Head	16	2	88.89%	9	1	90.00%
FHM	47	1	97.92%	19	0	100.00%
Head Instability	30	0	100.00%	16	0	100.00%
NMD	182	2	98.91%	146	0	100.00%
NPF	257	0	100.00%	182	0	100.00%
OTW	10	0	100.00%	6	0	100.00%
VDW	44	1	97.78%	45	0	100.00%
Weak Write	12	2	85.71%	8	0	100.00%
รวม	677	8	98.83%	475	1	99.79%

4.2.2 อัลกอริทึม Neural Network ผลการวิเคราะห์ข้อมูลการทำนายผลการเรียนรู้ด้วยข้อมูลสำหรับสอน (Training data) จำนวน 685 เรคคอร์ด มีความสามารถในการเรียนรู้ได้ถูกต้อง เป็นจำนวน 616 เรคคอร์ด คิดเป็น 89.93% และผลการเรียนรู้ไม่ถูกต้อง จำนวน 69 เรคคอร์ด คิดเป็น 10.07% และผลการทำนายด้วยข้อมูลทดสอบ (Testing Data) จำนวน 476 เรคคอร์ด มีความสามารถในการทำนายได้ถูกต้อง เป็นจำนวน 438 เรคคอร์ด คิดเป็น 92.02% และผลการทำนายไม่ถูกต้อง จำนวน 38 เรคคอร์ด คิดเป็น 7.98% จากข้อมูลทั้งหมด สามารถอธิบายผลที่เป็นรูปแบบ Confusion Matrix ดังแสดงในรูปที่ 4.19 และแสดงค่าความถูกต้องของแต่ละสาเหตุการเสียของแบบจำลองได้ตามตารางที่ 4.3

Results for output field Failure Mode

Comparing \$N\$-Failure Mode with Failure Mode

'Partition'	1_Training	2_Testing
Correct	616 89.93%	438 92.02%
Wrong	69 10.07%	38 7.98%
Total	685	476

Coincidence Matrix for \$N\$-Failure Mode (rows show actuals)

'Partition' = 1_Training	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	2	0	0	0	0	1	0	0	1	1
CND	0	67	1	1	0	1	1	0	1	2
Degraded Head	1	2	10	0	0	1	2	0	2	0
FHM	0	2	3	37	2	1	3	0	0	0
Head Instability	0	2	0	0	22	3	2	0	1	0
NMD	0	0	0	1	0	180	3	0	0	0
NPF	0	0	1	0	0	1	254	0	1	0
OTW	1	0	1	0	0	0	1	5	2	0
VDW	0	1	0	1	1	3	2	0	36	1
Weak write	0	1	0	1	1	4	1	0	3	3
'Partition' = 2_Testing	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	2	0	0	1	0	0	0	0	1	0
CND	0	30	0	0	1	5	2	0	1	1
Degraded Head	0	0	4	0	0	1	3	0	2	0
FHM	1	0	0	18	0	0	0	0	0	0
Head Instability	0	2	0	0	13	0	0	0	1	0
NMD	0	0	0	0	0	142	3	0	1	0
NPF	0	1	1	0	0	2	178	0	0	0
OTW	0	0	1	0	0	1	0	4	0	0
VDW	0	0	0	0	0	2	2	0	41	0
Weak write	0	0	0	0	0	2	0	0	0	6

รูปที่ 4.19 ผลการทดสอบของแบบจำลอง Neural Network

ตารางที่ 4.3 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของแบบจำลอง Neural Network

Failure Mode	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
ATI/STE	2	3	40.00%	2	2	50.00%
CND	67	7	90.54%	30	10	75.00%
Degraded Head	10	8	55.56%	4	6	40.00%
FHM	37	11	77.08%	18	1	94.74%
Head Instability	22	8	73.33%	13	3	81.25%
NMD	180	4	97.83%	142	4	97.26%
NPF	254	3	98.83%	178	4	97.80%
OTW	5	5	50.00%	4	2	66.67%
VDW	36	9	80.00%	41	4	91.11%
Weak Write	3	11	21.43%	6	2	75.00%
รวม	616	69	89.93%	438	38	92.02%



4.2.3 อัลกอริทึม C&R Tree ผลการวิเคราะห์ข้อมูลการทำนายผลการเรียนด้วยข้อมูลสำหรับสอน (Training data) จำนวน 685 เรคคอร์ด มีความสามารถในการเรียนรู้ได้ถูกต้อง เป็นจำนวน 354 เรคคอร์ด คิดเป็น 51.68% และผลการเรียนรู้ไม่ถูกต้อง จำนวน 331 เรคคอร์ด คิดเป็น 48.32% และการทำนายด้วยข้อมูลทดสอบ (Testing Data) จำนวน 476 เรคคอร์ด มีความสามารถในการทำนายได้ถูกต้อง เป็นจำนวน 272 เรคคอร์ด คิดเป็น 57.14% และการทำนายไม่ถูกต้อง จำนวน 204 เรคคอร์ด คิดเป็น 42.86% จากข้อมูลทั้งหมด สามารถอธิบายผลที่เป็นรูปแบบ Confusion Matrix ดังแสดงในรูปที่ 4.20 และแสดงค่าความถูกต้องของแต่ละสาเหตุการเสียของแบบจำลองได้ตามตารางที่ 4.4

Results for output field Failure Mode

Comparing \$R-Failure Mode with Failure Mode

'Partition'	1_Training		2_Testing	
Correct	354	51.68%	272	57.14%
Wrong	331	48.32%	204	42.86%
Total	685		476	

Coincidence Matrix for \$R-Failure Mode (rows show actuals)

'Partition' = 1_Training	CND	FHM	NMD	NPF	VDW
ATI/STE	0	0	2	3	0
CND	11	0	37	23	3
Degraded Head	1	0	10	6	1
FHM	1	3	27	17	0
Head Instability	1	0	22	7	0
NMD	1	3	123	56	1
NPF	0	1	45	211	0
OTW	0	0	6	3	1
VDW	1	1	23	14	6
Weak write	0	0	10	4	0

'Partition' = 2_Testing	CND	FHM	NMD	NPF	VDW
ATI/STE	1	0	1	2	0
CND	5	0	19	16	0
Degraded Head	0	0	5	3	2
FHM	0	1	11	6	1
Head Instability	1	0	9	5	1
NMD	1	1	99	44	1
NPF	0	1	25	156	0
OTW	0	0	5	1	0
VDW	1	0	21	12	11
Weak write	0	1	4	3	0

รูปที่ 4.20 ผลการทดสอบของแบบจำลอง C&R Tree

ตารางที่ 4.4 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของแบบจำลอง C&R Tree

Failure Mode	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
ATI/STE	0	5	0.00%	0	4	0.00%
CND	11	63	14.86%	5	35	12.50%
Degraded Head	0	18	0.00%	0	10	0.00%
FHM	3	45	6.25%	1	18	5.26%
Head Instability	0	30	0.00%	0	16	0.00%
NMD	123	61	66.85%	99	47	67.81%
NPF	211	46	82.10%	156	26	85.71%
OTW	0	10	0.00%	0	6	0.00%
VDW	6	39	13.33%	11	34	24.44%
Weak Write	0	14	0.00%	0	8	0.00%
รวม	354	331	51.68%	272	204	57.14%

4.2.4 อัลกอริทึม SVM ผลการวิเคราะห์ข้อมูลการทำนายผลการเรียนด้วยข้อมูลสำหรับสอน (Training data) จำนวน 685 เรคคอร์ด มีความสามารถในการเรียนรู้ได้ถูกต้อง เป็นจำนวน 480 เรคคอร์ด คิดเป็น 70.07% และผลการเรียนรู้ไม่ถูกต้อง จำนวน 205 เรคคอร์ด คิดเป็น 29.93% และผลการทำนายด้วยข้อมูลทดสอบ (Testing Data) จำนวน 476 เรคคอร์ด มีความสามารถในการทำนายได้ถูกต้อง เป็นจำนวน 340 เรคคอร์ด คิดเป็น 71.43% และผลการทำนายไม่ถูกต้อง จำนวน 136 เรคคอร์ด คิดเป็น 28.57% จากข้อมูลทั้งหมด สามารถอธิบายผลที่เป็นรูปแบบ Confusion Matrix ดังแสดงในรูปที่ 4.21 และแสดงค่าความถูกต้องของแต่ละสาเหตุการเสียของแบบจำลองได้ตามตารางที่ 4.5

Results for output field Failure Mode

Comparing \$S-Failure Mode with Failure Mode

'Partition'	1_Training	2_Testing
Correct	480 70.07%	340 71.43%
Wrong	205 29.93%	136 28.57%
Total	685	476

Coincidence Matrix for \$S-Failure Mode (rows show actuals)

'Partition' = 1_Training	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	2	0	0	0	0	2	1	0	0	0
CND	0	39	1	0	1	19	13	0	1	0
Degraded Head	0	0	12	0	0	2	2	0	2	0
FHM	0	2	1	19	0	11	12	1	2	0
Head Instability	0	0	0	0	14	10	5	0	1	0
NMD	0	4	1	2	0	137	40	0	0	0
NPF	0	6	0	1	0	18	231	0	1	0
OTW	0	0	0	0	0	4	1	5	0	0
VDW	0	2	2	2	1	12	8	0	18	0
Weak write	0	0	0	0	0	5	4	0	2	3

'Partition' = 2_Testing	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	2	0	0	0	0	1	1	0	0	0
CND	0	16	0	0	0	15	8	0	1	0
Degraded Head	0	0	4	1	0	2	2	0	1	0
FHM	0	0	0	11	0	2	6	0	0	0
Head Instability	0	1	0	0	8	4	2	0	1	0
NMD	0	5	0	0	0	106	30	0	5	0
NPF	0	2	0	1	0	15	163	0	1	0
OTW	0	0	0	0	0	2	0	4	0	0
VDW	0	1	0	1	0	12	9	0	22	0
Weak write	0	1	0	0	0	2	1	0	0	4

รูปที่ 4.21 ผลการทดสอบของแบบจำลอง SVM

ตารางที่ 4.5 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของแบบจำลอง SVM

Failure Mode	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
ATI/STE	2	3	40.00%	2	2	50.00%
CND	39	35	52.70%	16	24	40.00%
Degraded Head	12	6	66.67%	4	6	40.00%
FHM	19	29	39.58%	11	8	57.89%
Head Instability	14	16	46.67%	8	8	50.00%
NMD	137	47	74.46%	106	40	72.60%
NPF	231	26	89.88%	163	19	89.56%
OTW	5	5	50.00%	4	2	66.67%
VDW	18	27	40.00%	22	23	48.89%
Weak Write	3	11	21.43%	4	4	50.00%
รวม	480	205	70.07%	340	136	71.43%

4.2.5 อัลกอริทึม CHAID ผลการวิเคราะห์ข้อมูลการทำนายผลการเรียนด้วยข้อมูลสำหรับสอน (Training data) จำนวน 685 เรคคอร์ด มีความสามารถในการเรียนรู้ได้ถูกต้อง เป็นจำนวน 368 เรคคอร์ด คิดเป็น 53.72% และผลการเรียนรู้ไม่ถูกต้อง จำนวน 317 เรคคอร์ด คิดเป็น 46.28% และผลการทำนายด้วยข้อมูลทดสอบ (Testing Data) จำนวน 476 เรคคอร์ด มีความสามารถในการทำนายได้ถูกต้อง เป็นจำนวน 243 เรคคอร์ด คิดเป็น 51.05% และผลการทำนายไม่ถูกต้อง จำนวน 233 เรคคอร์ด คิดเป็น 48.95% จากข้อมูลทั้งหมด สามารถอธิบายผลที่เป็นรูปแบบ Confusion Matrix ดังแสดงในรูปที่ 4.22 และแสดงค่าความถูกต้องของแต่ละสาเหตุการเสียของแบบจำลองได้ตามตารางที่ 4.6

Results for output field Failure Mode

Comparing \$R-Failure Mode with Failure Mode

'Partition'	1_Training		2_Testing	
Correct	368	53.72%	243	51.05%
Wrong	317	46.28%	233	48.95%
Total	685		476	

Coincidence Matrix for \$R-Failure Mode (rows show actuals)

'Partition' = 1_Training	CND	Degraded Head	FHM	Head Instability	NMD	NPF
AT/STE	1	0	0	0	1	3
CND	19	6	0	1	29	19
Degraded Head	0	9	0	0	3	6
FHM	5	3	3	2	20	15
Head Instability	4	1	1	6	9	9
NMD	4	1	3	2	100	74
NPF	2	0	0	1	23	231
OTW	0	1	0	1	5	3
VDW	6	5	0	0	20	14
Weak write	0	0	1	0	7	6
'Partition' = 2_Testing	CND	Degraded Head	FHM	Head Instability	NMD	NPF
AT/STE	0	0	0	0	2	2
CND	11	2	1	0	14	12
Degraded Head	1	1	0	1	1	6
FHM	3	0	2	2	8	4
Head Instability	3	1	0	2	3	7
NMD	8	6	3	3	63	63
NPF	1	0	0	0	17	164
OTW	0	0	0	2	2	2
VDW	4	3	0	0	23	15
Weak write	0	0	2	0	3	3

รูปที่ 4.22 ผลการทดสอบของแบบจำลอง CHAID

ตารางที่ 4.6 ผลการวิเคราะห์ข้อมูลที่ใช้สำหรับสอนและข้อมูลที่ใช้ทดสอบของแบบจำลอง CHAID

Failure Mode	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
ATI/STE	0	5	0.00%	0	4	0.00%
CND	19	55	25.68%	11	29	27.50%
Degraded Head	9	9	50.00%	1	9	10.00%
FHM	3	45	6.25%	2	17	10.53%
Head Instability	6	24	20.00%	2	14	12.50%
NMD	100	84	54.35%	63	83	43.15%
NPF	231	26	89.88%	164	18	90.11%
OTW	0	10	0.00%	0	6	0.00%
VDW	0	45	0.00%	0	45	0.00%
Weak Write	0	14	0.00%	0	8	0.00%
รวม	368	317	53.72%	243	233	51.05%

### 4.3 สรุปผลการทดลอง

จากผลดำเนินการสร้างแบบจำลองจากอัลกอริทึม C5.0, Neural Network, C&R Tree, SVM และ CHAID เมื่อเปรียบเทียบประสิทธิภาพพบว่าการคัดแยกสาเหตุการเสียบของฮาร์ดดิสก์ด้วยแบบจำลอง C5.0 มีความถูกต้องมากที่สุด ดังตารางที่ 4.7 คือ แบบจำลอง C5.0 สามารถคัดแยกได้ถูกต้องเป็นจำนวน 475 เรคคอร์ด คิดเป็น 99.79% แบบจำลอง Neural Network สามารถคัดแยกได้ถูกต้องเป็นจำนวน 438 เรคคอร์ด คิดเป็น 92.02% แบบจำลอง C&R Tree สามารถคัดแยกได้ถูกต้องเป็นจำนวน 272 เรคคอร์ด คิดเป็น 57.14% แบบจำลอง SVM สามารถคัดแยกได้ถูกต้องเป็นจำนวน 340 เรคคอร์ด คิดเป็น 71.43% และแบบจำลอง CHAID สามารถคัดแยกได้ถูกต้องเป็นจำนวน 243 เรคคอร์ด คิดเป็น 51.05% ค่าความถูกต้องของแต่ละแบบจำลองมีค่าที่แตกต่างกันมากเนื่องจากแต่ละอัลกอริทึมมีความสามารถในการคัดแยกข้อมูลที่แตกต่างกัน ทั้งนี้ค่าความถูกต้องที่ได้ขึ้นอยู่กับการแบ่งสัดส่วนข้อมูลที่คัดแยก และข้อมูลที่เลือกมาด้วย เพราะบางอัลกอริทึมเหมาะสมกับการคัดแยกข้อมูลที่มีจำนวนกลุ่มข้อมูลมาก ๆ ได้และมีความสามารถในการประมวลผลข้อมูลหลายตัวอย่างได้ดี แต่บางอัลกอริทึมสามารถให้ผลลัพธ์ในการคัดแยกได้ดีกับกลุ่มข้อมูลที่มีจำนวนน้อย

ตารางที่ 4.7 ค่าความถูกต้องโดยรวมของแต่ละแบบจำลอง

Algorithm	Training data			Testing data		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
C5.0	677	8	98.83%	475	1	99.79%
Neural Network	616	69	89.93%	438	38	92.02%
C&R Tree	354	331	51.68%	272	204	57.14%
SVM	480	205	70.07%	340	136	71.43%
CHAID	368	317	53.72%	243	233	51.05%

เมื่อทำการทดลองที่ได้ผ่านกระบวนการหลายขั้นตอน ประกอบด้วยการประมวลผล และการวิเคราะห์ทำให้ได้คำตอบแต่ละอัลกอริทึมซึ่งเป็นไปตามสมมติฐาน ทำให้สามารถนำไปสู่การสรุปผลการวิจัยต่อไป



## บทที่ 5

### สรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะ

จากที่ได้ทราบถึงวิธีดำเนินงานวิจัยและผลการทดลองแล้ว ในบทนี้จะนำเสนอข้อสรุปงานวิจัยทั้งหมด ข้อเสนอแนะ รวมถึงปัญหาที่พบในงานวิจัยซึ่งอาจเป็นประโยชน์ต่อผู้ที่สนใจ โดยมีรายละเอียดดังต่อไปนี้

#### 5.1 สรุปผลการวิจัยและการอภิปรายผล

งานวิจัยนี้จึงได้มีการใช้อัลกอริทึมต่าง ๆ มาสร้างแบบจำลองโดยให้เรียนรู้ประเภทการเสียของฮาร์ดดิสก์จากการเลือกข้อมูลแอตทริบิวต์ในกระบวนการทดสอบฮาร์ดดิสก์มาปรับแต่งข้อมูลให้สอดคล้องกับการทำงานของแต่ละอัลกอริทึม เมื่อนำผลลัพธ์ที่ได้มาเปรียบเทียบกับพบว่าแบบจำลองที่ได้จากอัลกอริทึม C5.0 สามารถคัดแยกสาเหตุการเสียของฮาร์ดดิสก์ได้ถูกต้องมากกว่าแบบจำลองที่ได้จากอัลกอริทึมอื่นซึ่งเป็นไปตามสมมติฐาน เนื่องจากอัลกอริทึม C5.0 มีความสามารถในการเรียนรู้รูปแบบของข้อมูลได้ดี สามารถประมวลผลได้กับข้อมูลที่มีขนาดใหญ่ และผลลัพธ์ที่ได้มีความถูกต้องแม่นยำสูงในขณะที่ใช้ระยะเวลาการประมวลผลน้อย เพราะมีการเพิ่มความสามารถจากการปรับค่าพารามิเตอร์ให้มีการใช้งาน Boosting สำหรับสร้างต้นไม้ตัดสินใจหลายต้นให้กับอัลกอริทึม ทำให้เพิ่มความสามารถในการประมวลผลอัลกอริทึม C5.0 ผลลัพธ์ที่ได้เป็นแบบจำลองต้นไม้ตัดสินใจที่มีต้นไม้ตัดสินใจหลายต้นมาช่วยกันคัดแยกสาเหตุการเสียของฮาร์ดดิสก์จากข้อมูลที่มีจำนวนมากได้ และมีการเพิ่มเทคนิคการตัดกิ่งที่ไม่จำเป็นออกเพื่อลดจำนวนแอตทริบิวต์ที่ไม่จำเป็นและช่วยลดระยะเวลาในการประมวลผลการทำงานอีกด้วย ซึ่งอัลกอริทึมอื่นไม่สามารถทำได้ แต่ผลลัพธ์ที่ได้สามารถเปลี่ยนแปลงไปถ้ามีรูปแบบของข้อมูลที่ต่างไป ทั้งนี้ต้องขึ้นอยู่กับรูปแบบของข้อมูลที่ใช้วิเคราะห์ด้วย

จากการสร้างแบบจำลองที่ได้จากอัลกอริทึม C5.0 มีความสอดคล้องกับลำดับความสำคัญของแอตทริบิวต์ที่ถูกเลือกโดยผู้เชี่ยวชาญจึงทำให้แบบจำลองที่ได้มีความน่าเชื่อถือ สามารถนำไปใช้งานได้จริง ซึ่งทำให้ลดขั้นตอนทดสอบความน่าเชื่อถือและวิเคราะห์สาเหตุการเสียของฮาร์ดดิสก์

## 5.2 ข้อเสนอแนะ

ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลที่มีจำนวนแอตทริบิวต์มาก ซึ่งอาจมีข้อมูลในรูปแบบที่ต่างกันมาก ควรประยุกต์ใช้วิธีการลดขนาด การรวมค่า การหาค่าเฉลี่ย การจัดการ และเทคนิคอื่น ๆ ในการสรุปค่าของแต่ละแอตทริบิวต์เพื่อลดระยะเวลาและปัญหาในการประมวลผล

## 5.3 ปัญหาและอุปสรรค

5.3.1 ในการรวบรวมข้อมูลสาเหตุการเสียชีวิตของฮาร์ดดิสก์แต่ละประเภทมีความแตกต่างกันมากพอสมควร อีกทั้งมีความหลากหลายของข้อมูลที่ได้จากกระบวนการทดสอบฮาร์ดดิสก์ จึงเป็นอุปสรรคที่สำคัญในการเลือกแอตทริบิวต์ที่มีความเหมาะสมมาใช้ในการสร้างแบบจำลอง

5.3.2 เครื่องมือที่ใช้ในการสร้างตัวแบบอาจสร้างกฎที่ไม่ครอบคลุม ต้องทำการตรวจสอบข้อมูลและปรับแอตทริบิวต์ให้เหมาะสม





## รายการอ้างอิง

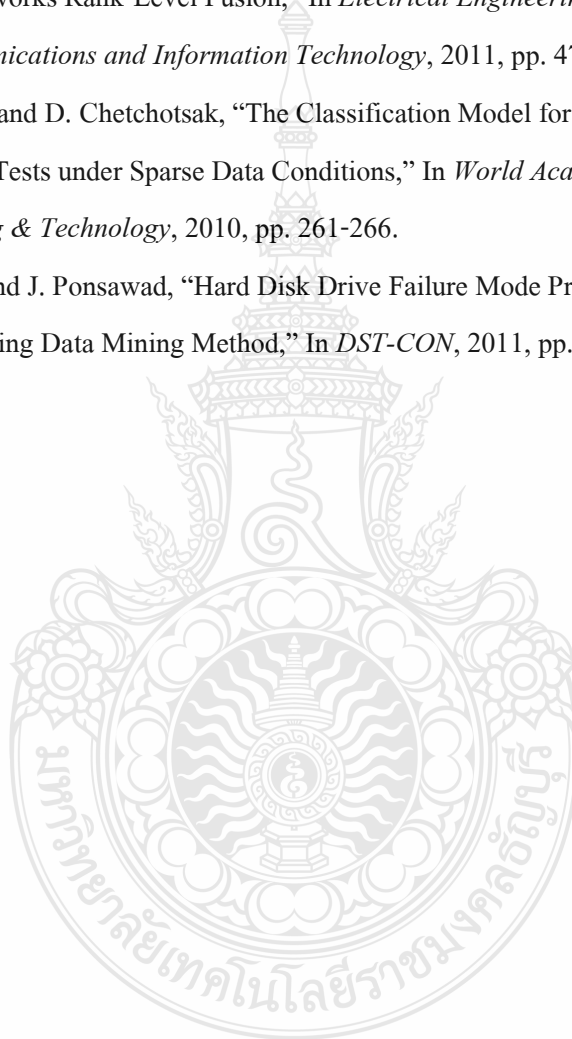
- [1] Hard disk drive reliability and MTBF/AFR (Online), Available: [http://knowledge.seagate.com/articles/en\\_US/FAQ/174791en?language=en\\_US](http://knowledge.seagate.com/articles/en_US/FAQ/174791en?language=en_US) (2556).
- [2] J.D. Ullman, *Principles of database and knowledge-base systems*, New York: Computer Science Press, 1988.
- [3] Abdullah Al Mamun, GuoXiao Guo, and Chao Bi, *Hard Disk Drive Mechatronics and Control*, New York: CRC Press, 2007.
- [4] Charles M. Kozierok, Construction and operation of the hard disk (Online), Available: <http://www.pcguide.com/ref/hdd/op/index.htm> (2554).
- [5] Seagate Technology LLC., Build Box-system, (2009).
- [6] B. D. Strom, S. C. Lee, G. W. Tyndall, and A. Khurshudov, "Hard disk drive reliability modeling and failure prediction," In *Asia-Pacific Magnetic Recording Conference*, 2006, pp. 1-2.
- [7] F.B. Sun, and S. Zhang, "Does Hard Disk Drive Failure Rate Enter Steady-State After One Year?" In *Reliability and Maintainability Symposium*, 2007, pp. 356-361.
- [8] Hongtao Chen, Chunqing Wang, and Mingyua Li, "Mechanical Shock Modeling and Testing of Lead-Free Solder Joint in Hard Disk Drive Head Assembly," In *Asian Green Electronics Conference*, 2005, pp. 192- 196.
- [9] R. Rai, and D. B. Bogy, "Parametric Study of Operational Shock in Mobile Disk Drives With Disk-Ramp Contact," *IEEE Transactions on Magnetics*, vol. 47, No. 7, pp. 1878-1881, 2011.
- [10] R. Rai, and D. B. Bogy, "Parametric study of HDD operational shock with disk -ramp contact," In *APMRC Digest*, 2010, pp. 1-2.
- [11] เรวดี ศักดิ์คุณขจรธรรม, "การใช้เทคนิคคาตาไมน์นิงในการสร้างฐานความรู้เพื่อการทำนายสัมฤทธิ์ผลทางการเรียนของนักศึกษาวิทยาลัยราชพฤกษ์," รายงานการวิจัย, วิทยาลัยราชพฤกษ์, 2553.
- [12] อุดุลย์ ยี่มงาม, การทำเหมืองข้อมูล (Online), Available: [http://compcenter.bu.ac.th/index.php?option=com\\_content&task=view&id=75&Itemid=172](http://compcenter.bu.ac.th/index.php?option=com_content&task=view&id=75&Itemid=172) (2551).

## รายการอ้างอิง (ต่อ)

- [13] ชินพัฒน แก้วชินพร, “การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจและการจัดกลุ่ม,” ปรินญาณิพนธ์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2553.
- [14] พวงทิพย์ แทนแสง, “การทดสอบประสิทธิภาพการทำงานของอัลกอริทึมการไมนิ่งกลูสำหรับจำแนก,” วิทยานิพนธ์ปรินญามหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2550.
- [15] S.A. Jafari, S. Mashohor, and M. Jalali Varnamkhasti, “Committee Neural Networks with Fuzzy Genetic Algorithm,” *Journal of Petroleum Science and Engineering*, vol. 76, No. 3-4, pp. 217-223, 2011.
- [16] วิวัฒน์ ชินนาทศิริกุล, การทำเหมืองข้อมูลแบบจำแนก (Online), Available: <http://www.csitvru.com/wiwat/mining/chapter7.ppt> (2556).
- [17] Niu Zhixian, Zong Lili, Yan Qingwei, and Zhao Zhenxing, “Auto-Recognizing Dbms Workload Based on C5.0 Algorithm,” In *Knowledge Discovery and Data Mining*, 2009, pp. 777-780.
- [18] อุบลวรรณ กิจคณะ, “การเรียนรู้รูปแบบบรรทัดฐานกรรมเพื่อจำแนกชนิดของโรคมะเร็งเม็ดเลือดขาวด้วยอัลกอริทึม C5.0,” วิทยานิพนธ์ปรินญามหาบัณฑิต มหาวิทยาลัยขอนแก่น, 2553.
- [19] ไพฑูรย์ จันทรเรือง, “ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคต้นไม้ตัดสินใจ,” วิทยานิพนธ์ปรินญามหาบัณฑิต, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2550.
- [20] นิตยา เกิดประสพ, “อัลกอริทึมและเทคนิคที่เหมาะสมกับการสังเคราะห์โมเดลที่ช่วยวินิจฉัยโรคได้อัตโนมัติ,” รายงานการวิจัย, มหาวิทยาลัยเทคโนโลยีสุรนารี, 2545.
- [21] Geoffrey I Webb, “MultiBoosting: A Technique for Combining Boosting and Wagging,” *Machine Learning*, vol. 40, No. 2, pp. 159-196, 2000.
- [22] Ian H. Witten, and Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, San Francisco: Morgan Kaufmann Publishers Inc., 1999.
- [23] Gutierrez-Osuna R., Cross-validation (Online), Available: [http://research.cs.tamu.edu/prism/lectures/pr/pr\\_113.pdf](http://research.cs.tamu.edu/prism/lectures/pr/pr_113.pdf) (2006).

## รายการอ้างอิง (ต่อ)

- [24] Howard Hamilton, Confusion Matrix (Online), Available: [http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion\\_matrix/confusion\\_matrix.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html) (2012).
- [25] W. Tepin, and Y. Kidjaidure, "Customer Failure Modes Prediction for Hard Disk Drives using Neural Networks Rank-Level Fusion," In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2011, pp. 476-479.
- [26] S. Pattanapairoj, and D. Chetchotsak, "The Classification Model for Hard Disk Drive Functional Tests under Sparse Data Conditions," In *World Academy of Science, Engineering & Technology*, 2010, pp. 261-266.
- [27] T. Ramangkul, and J. Ponsawad, "Hard Disk Drive Failure Mode Prediction from SMART Attribute using Data Mining Method," In *DST-CON*, 2011, pp. 338-340.

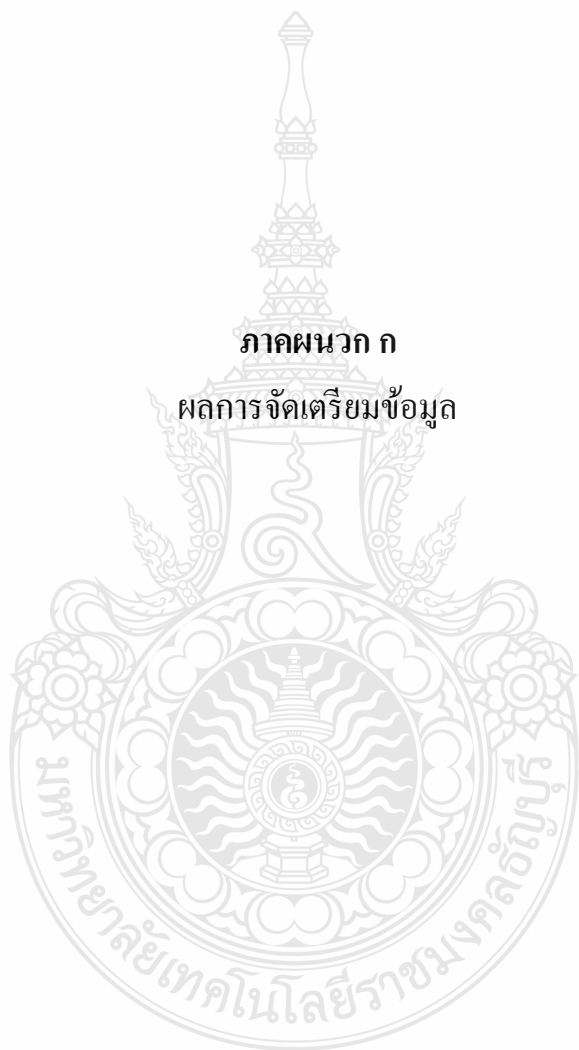


ภาคผนวก

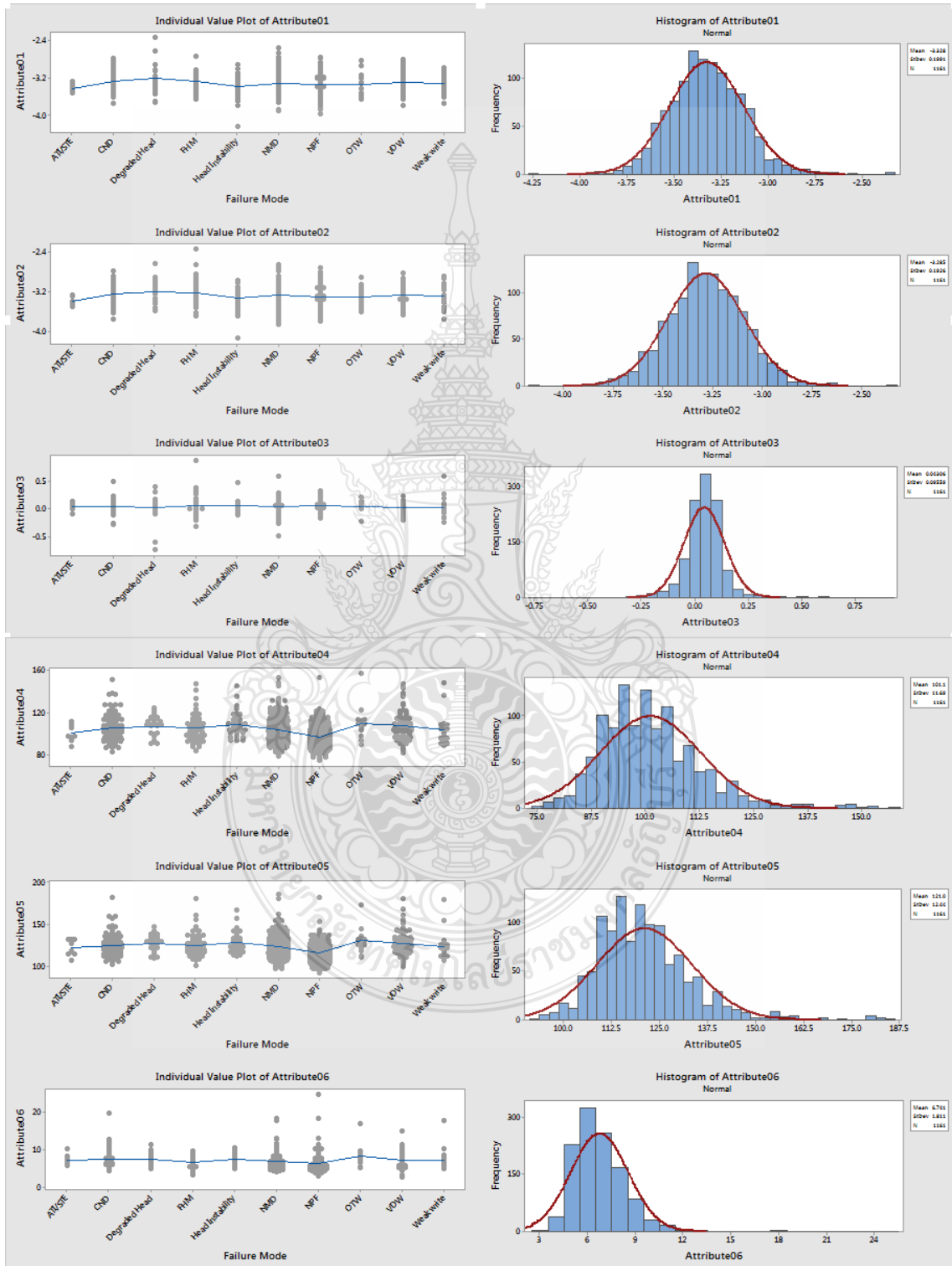


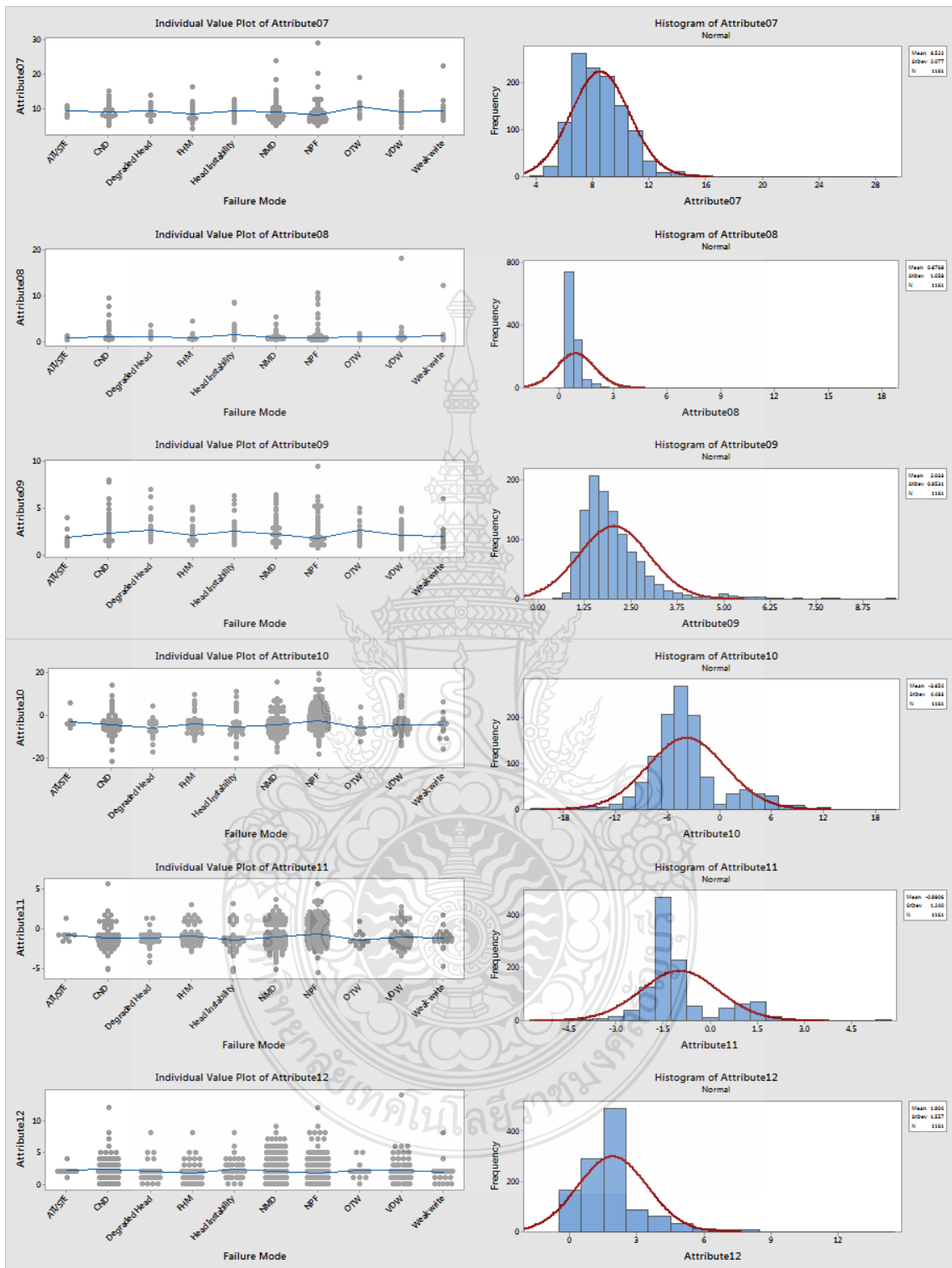
ภาคผนวก ก

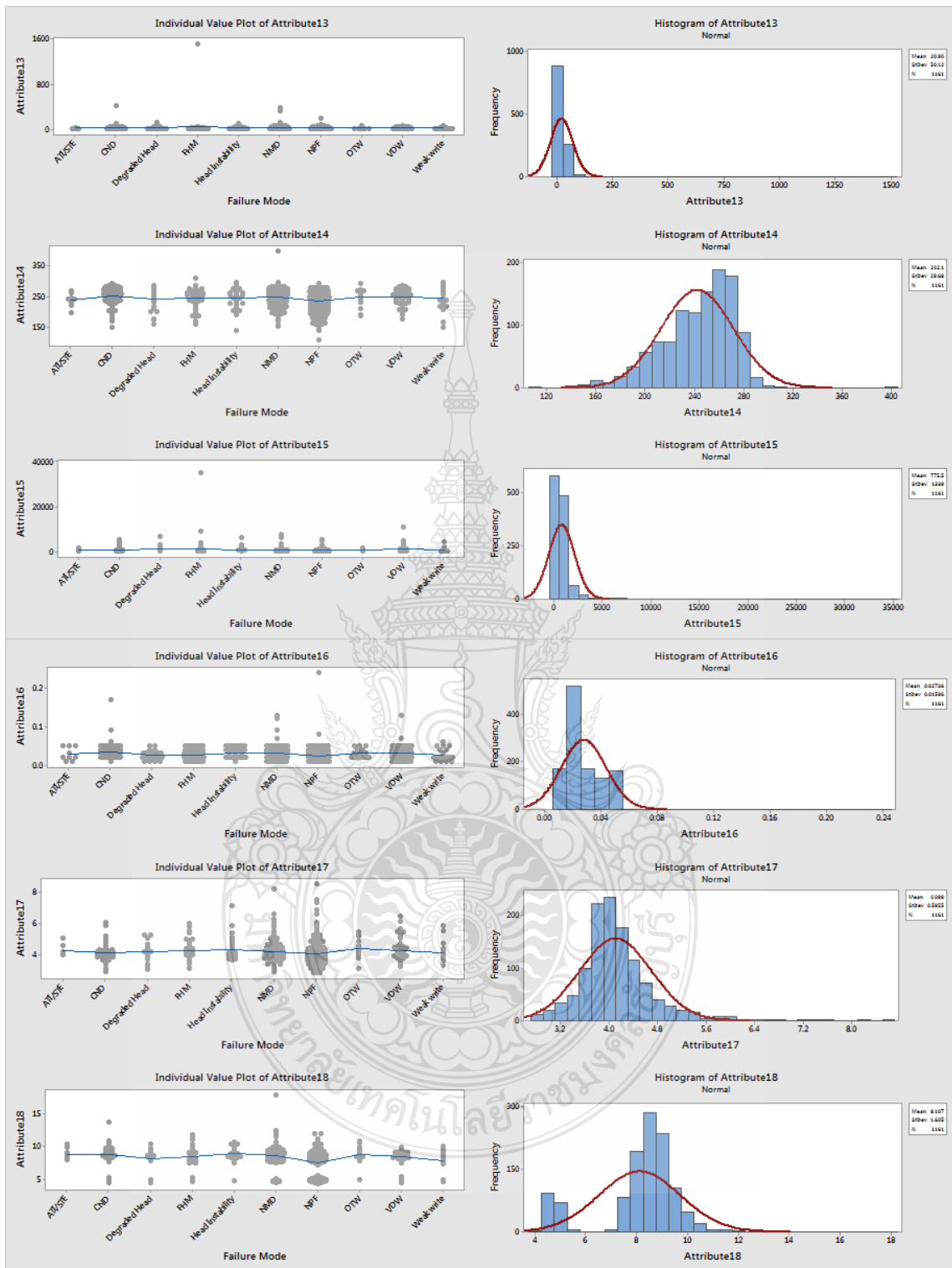
ผลการจัดเตรียมข้อมูล



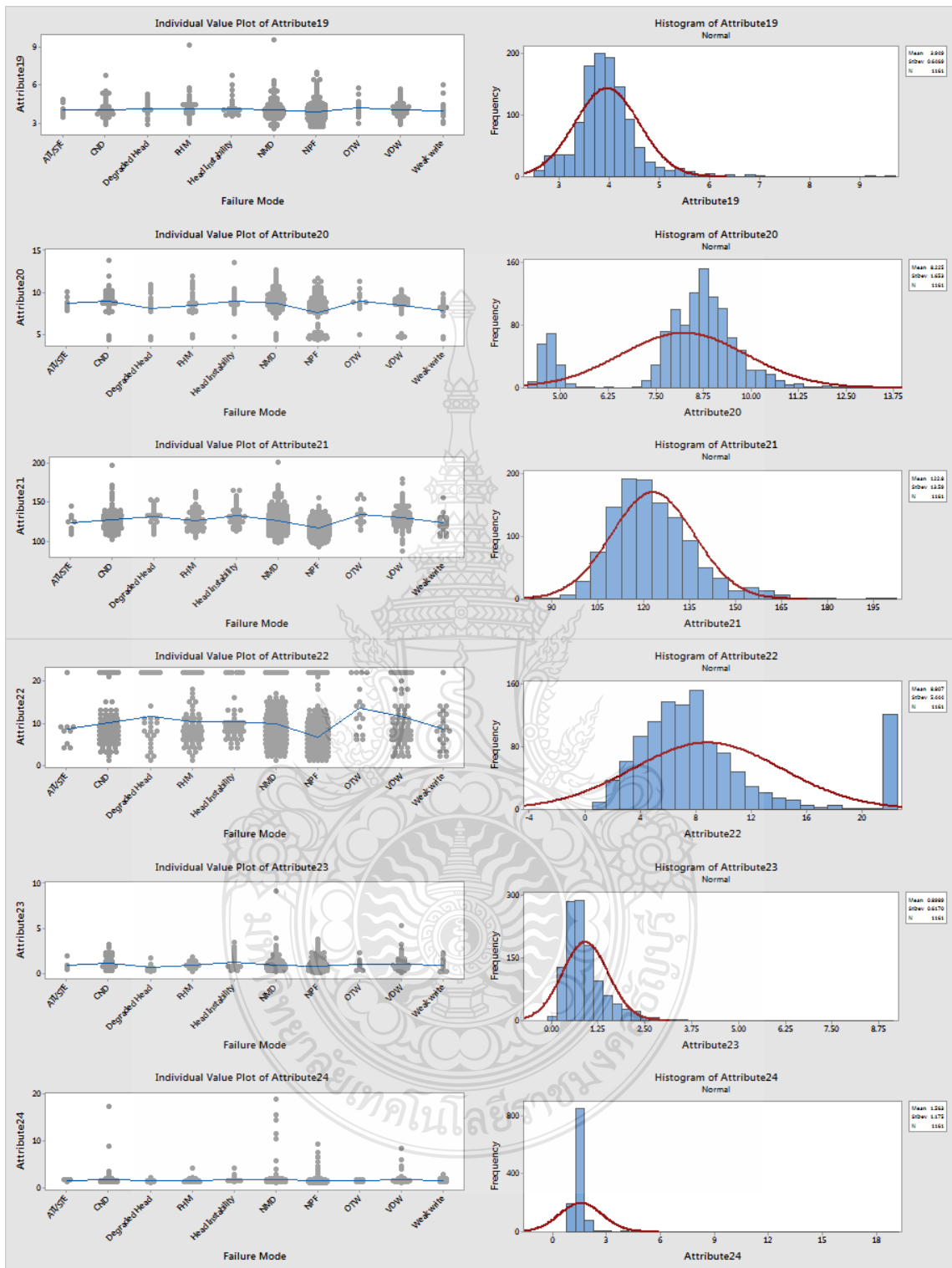
ผลการจัดเตรียมข้อมูลในแต่ละแอตทริบิวท์สามารถใช้กราฟ Individual value plot และ Histogram เพื่อดูลักษณะการกระจายของข้อมูลที่น่ามาทดสอบ โดยแยกตาม Failure Mode และ จำนวนฮาร์ดดิสก์ที่เกี่ยวข้อง

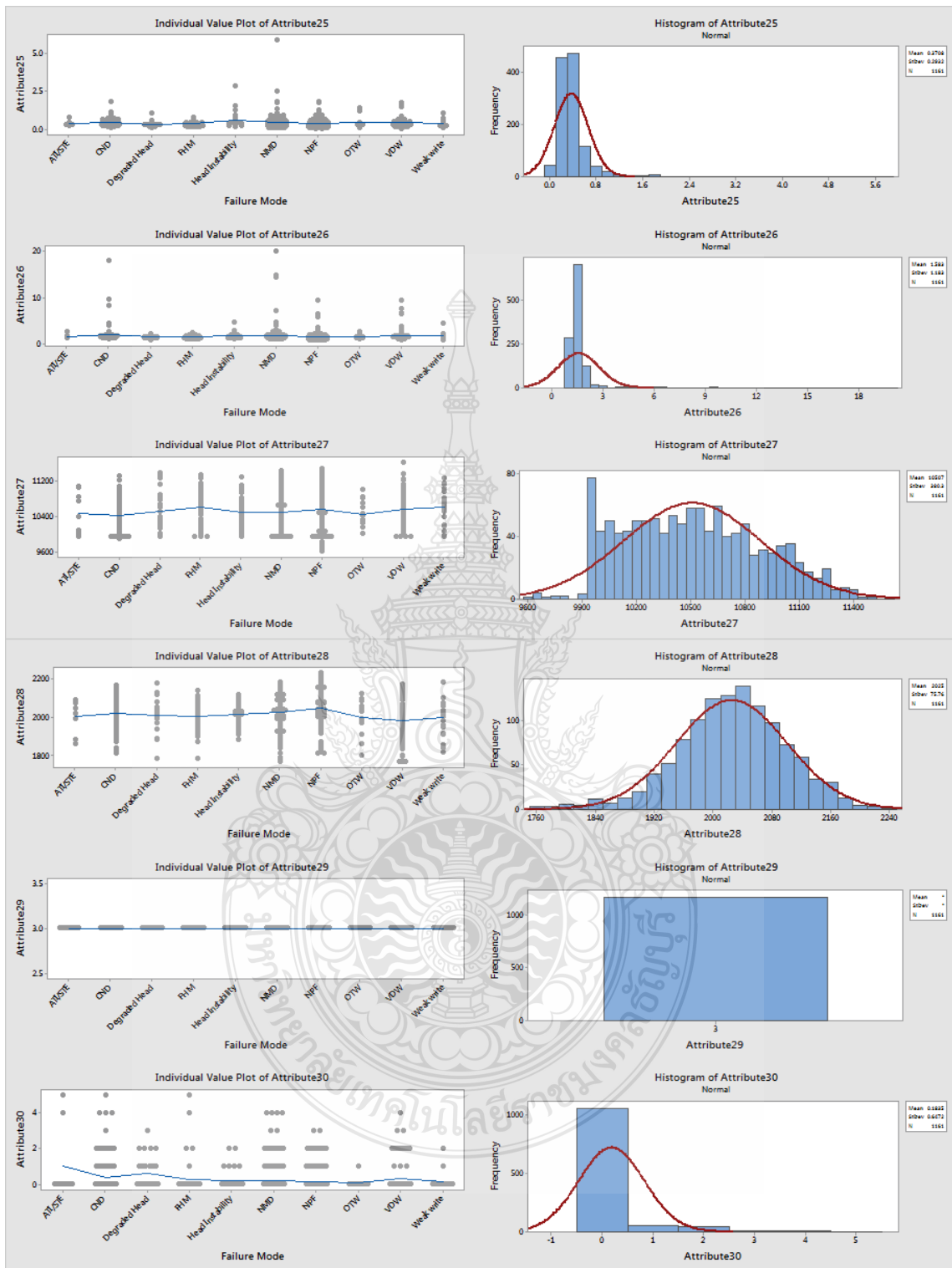


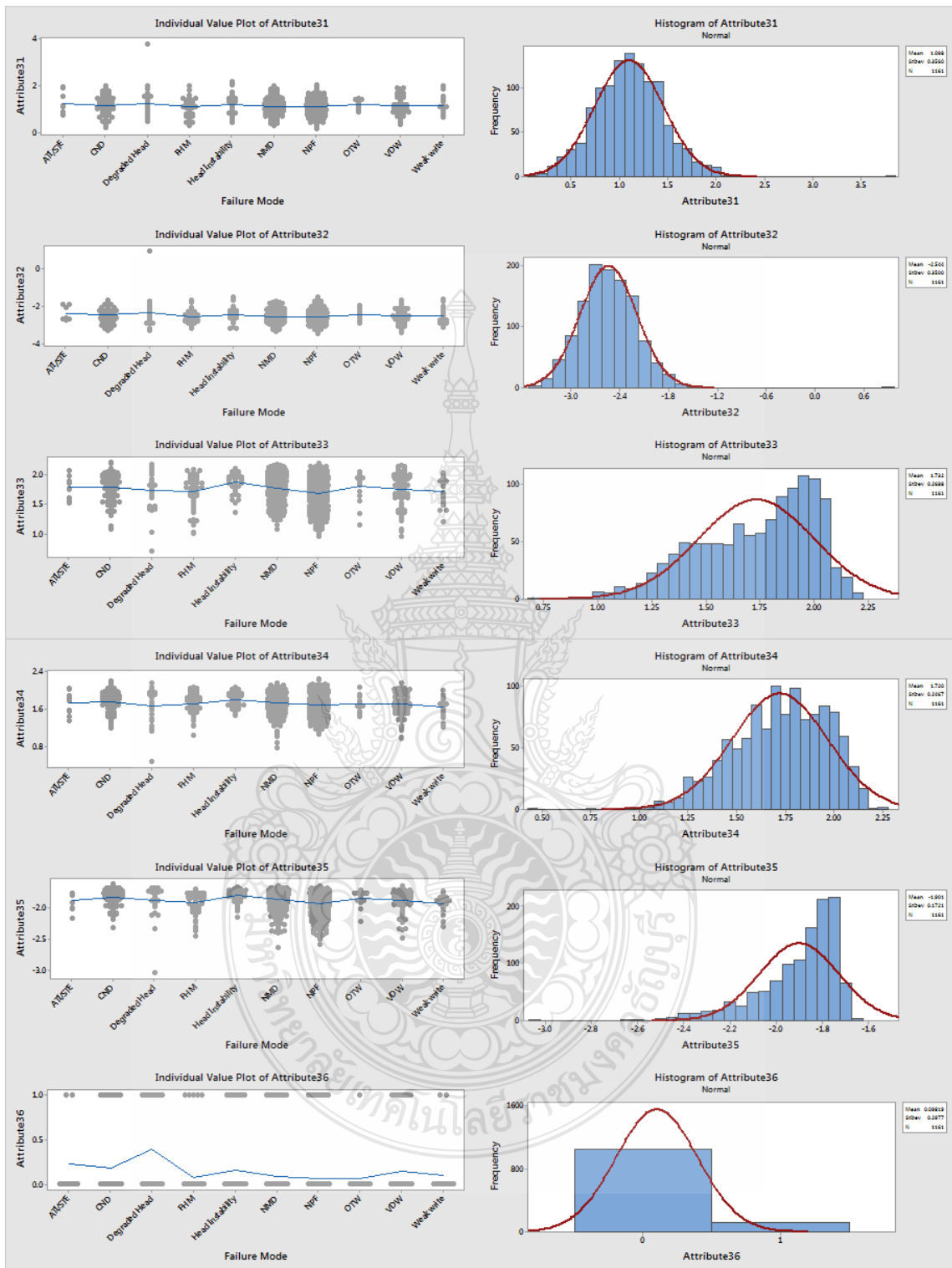


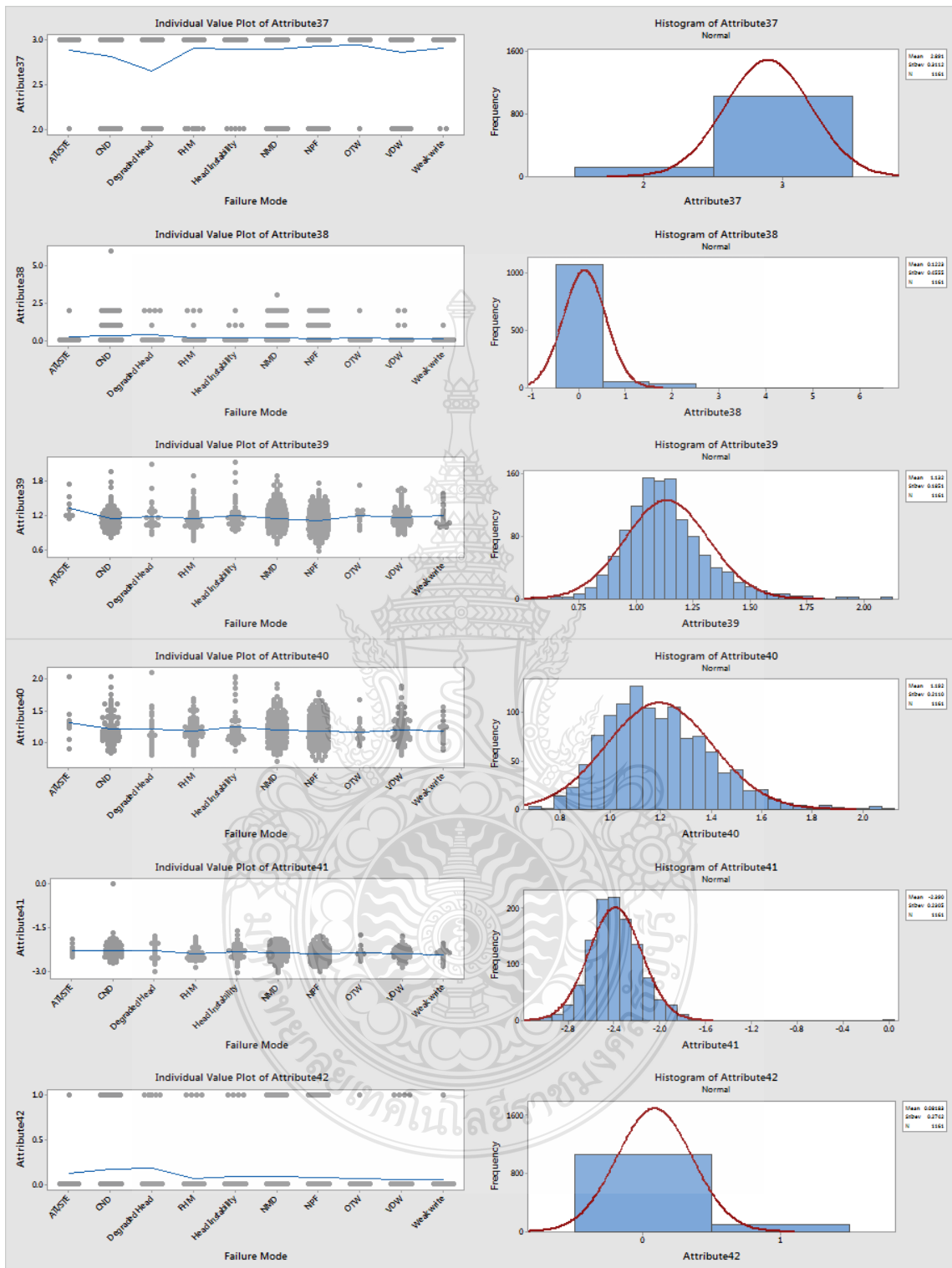


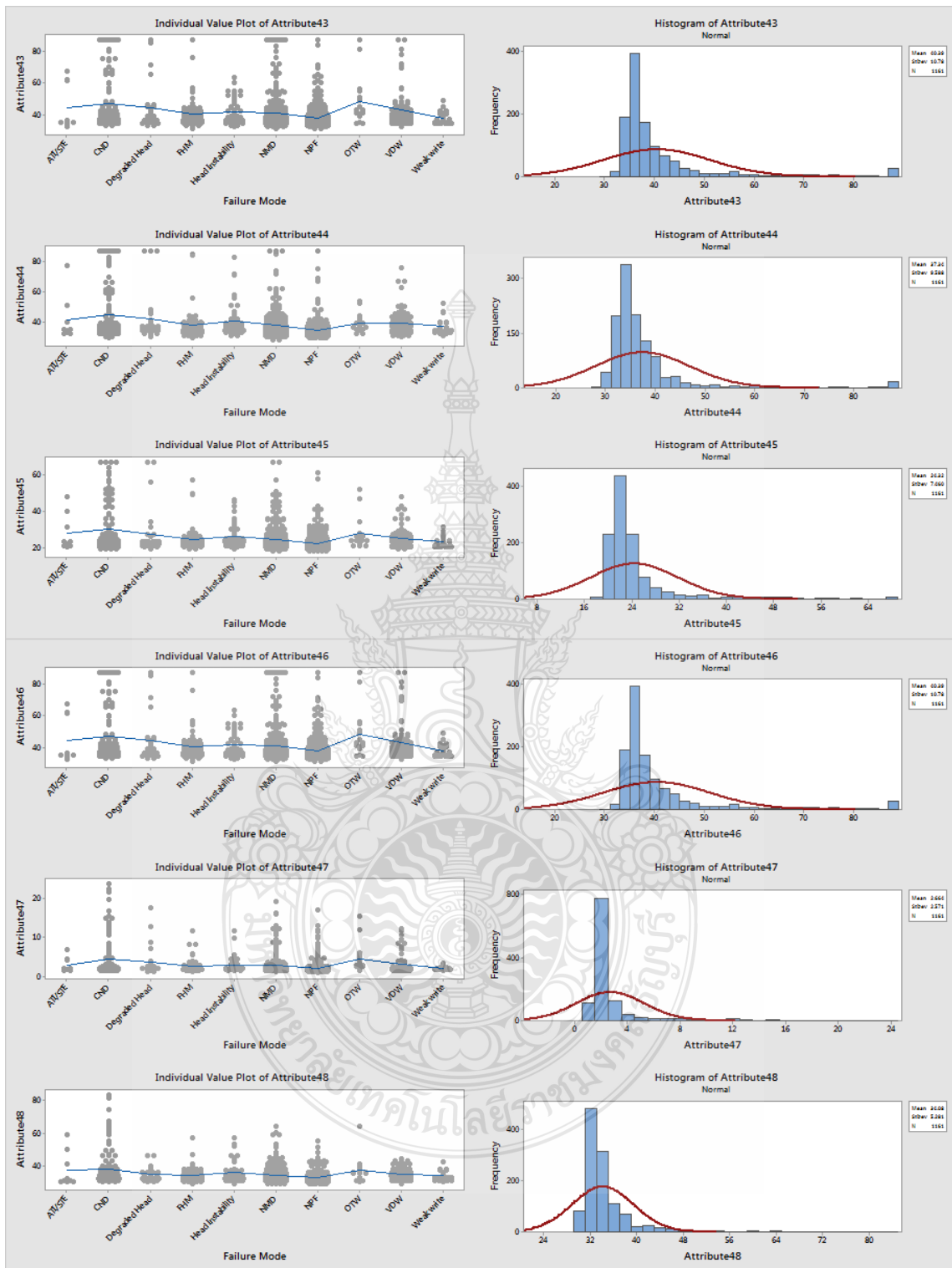


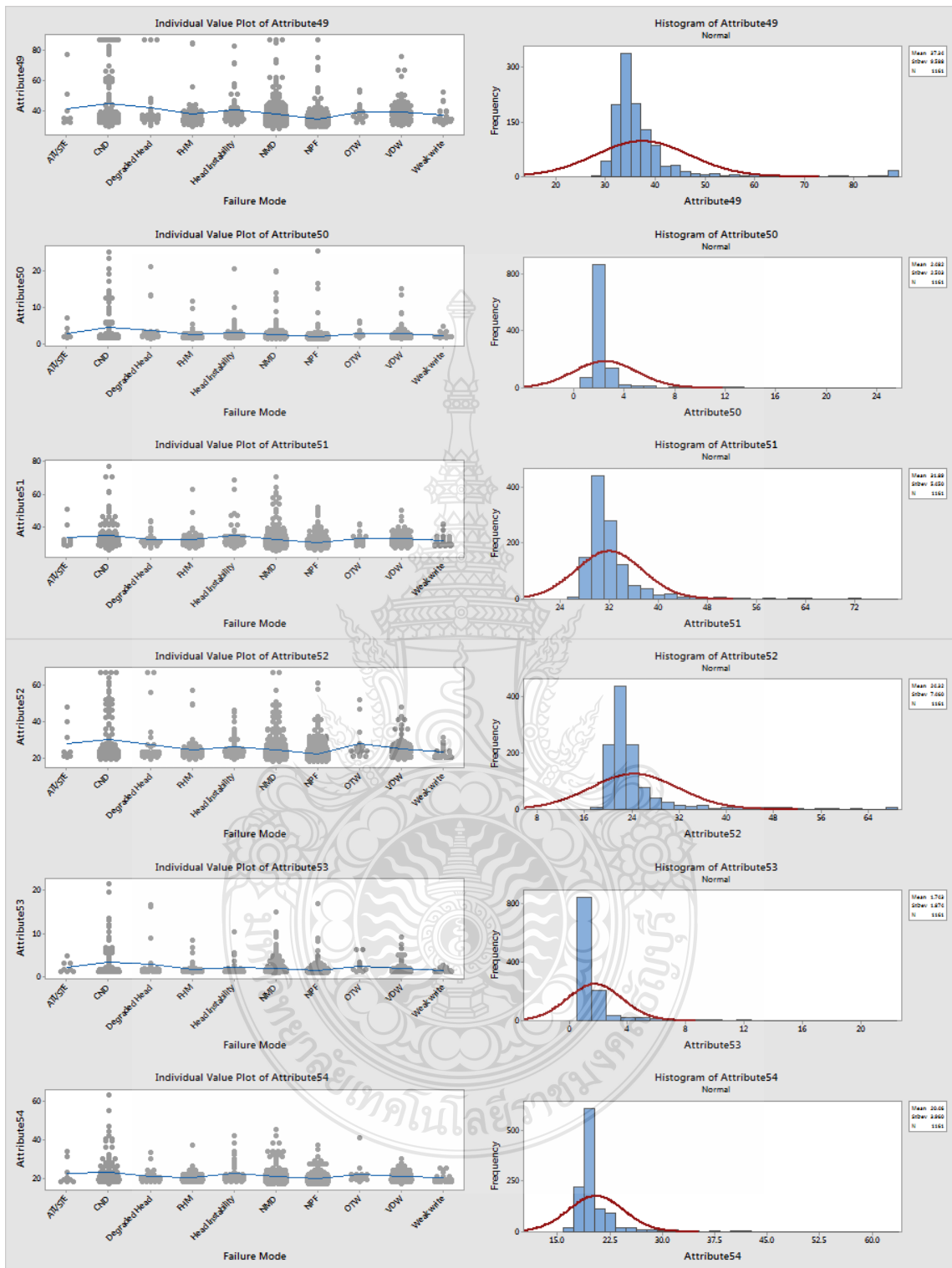


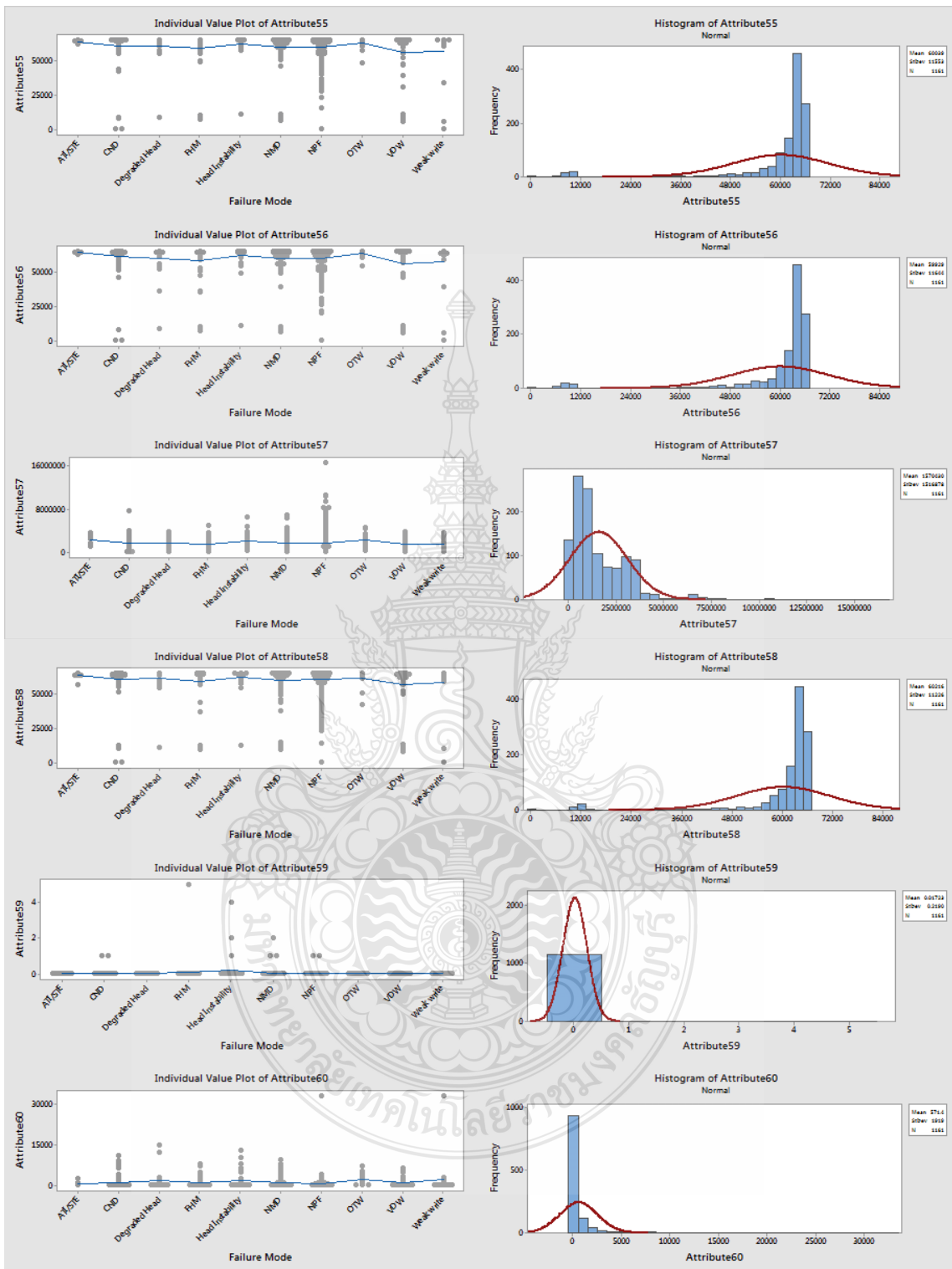


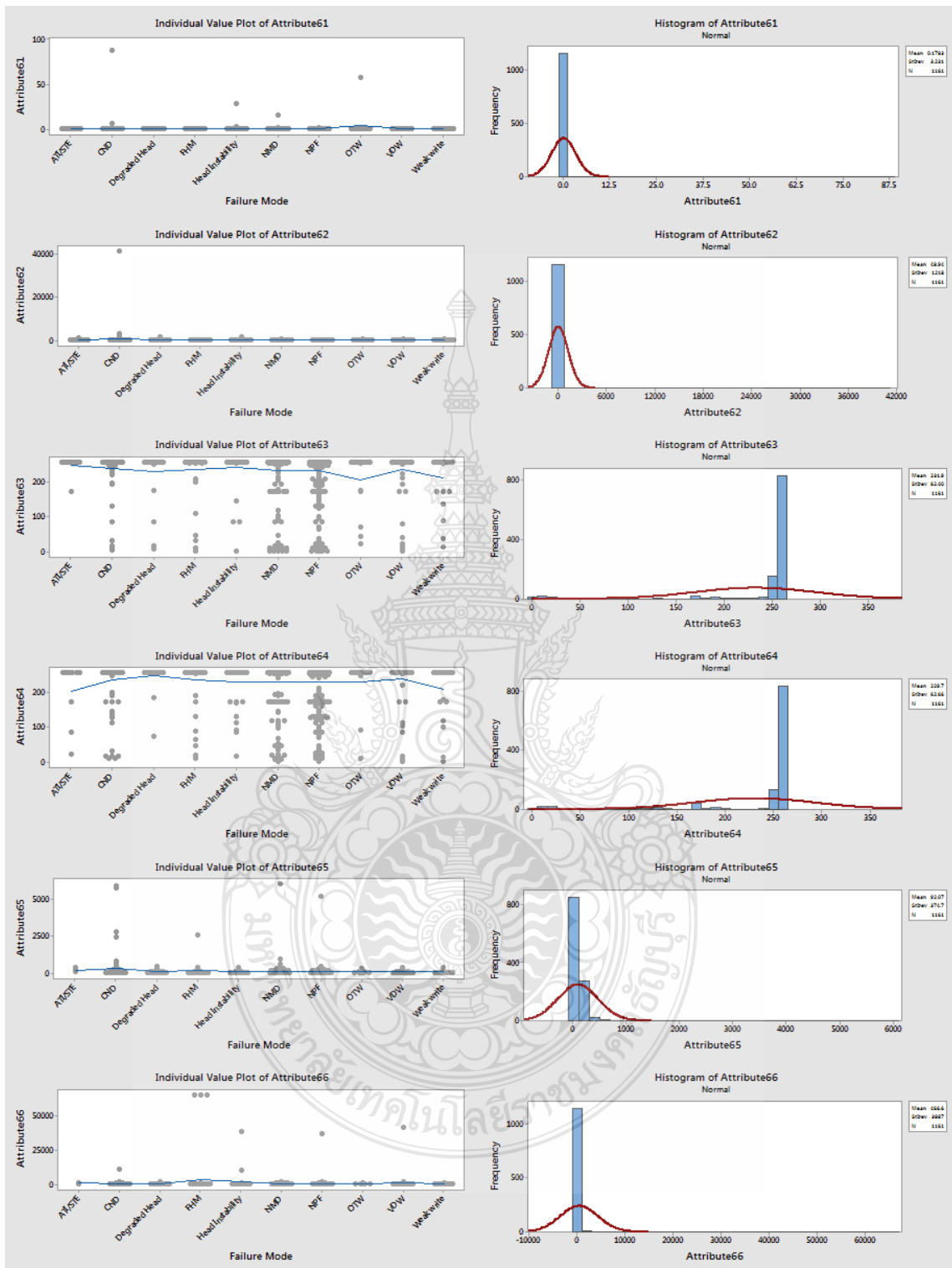




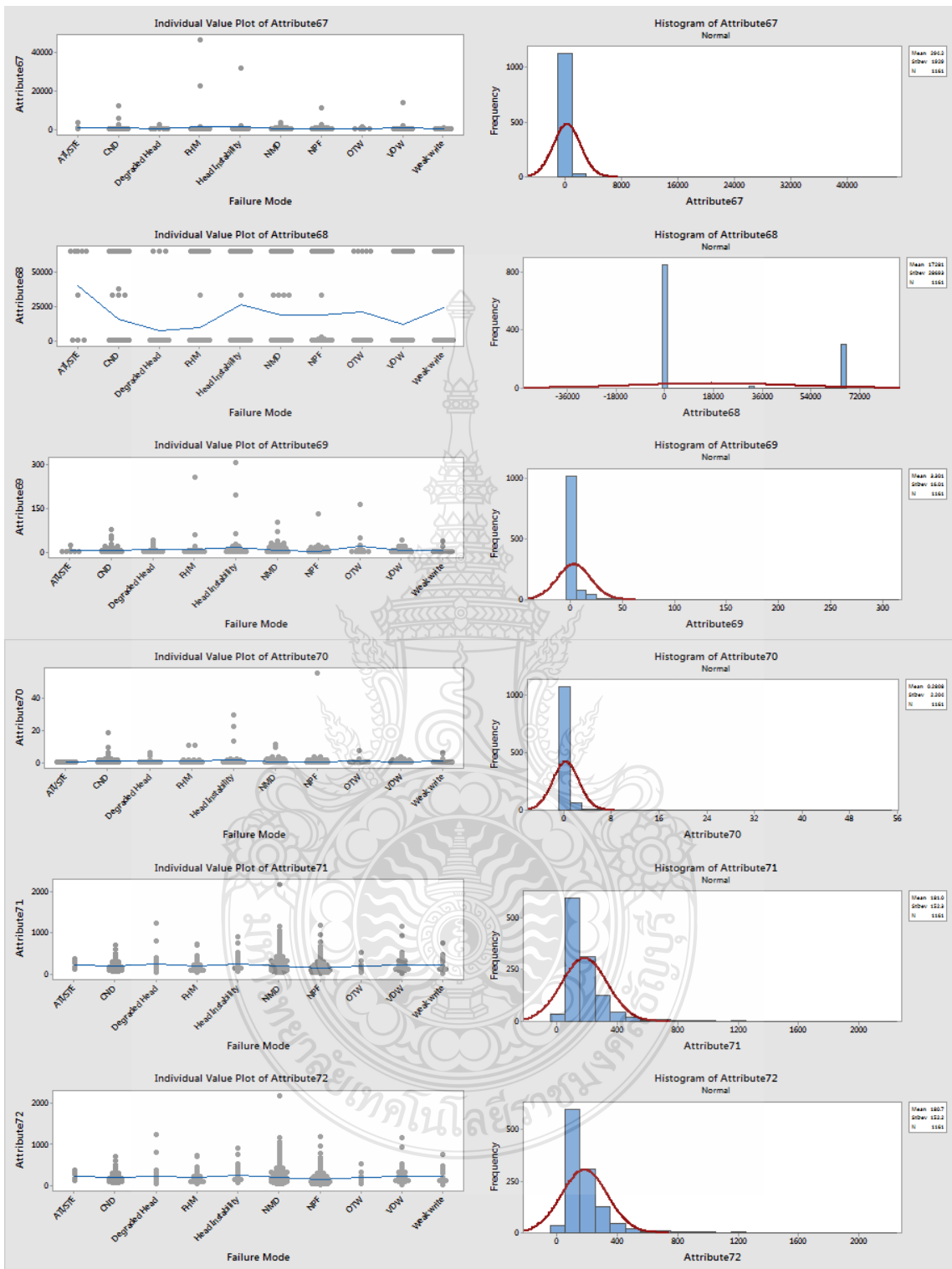


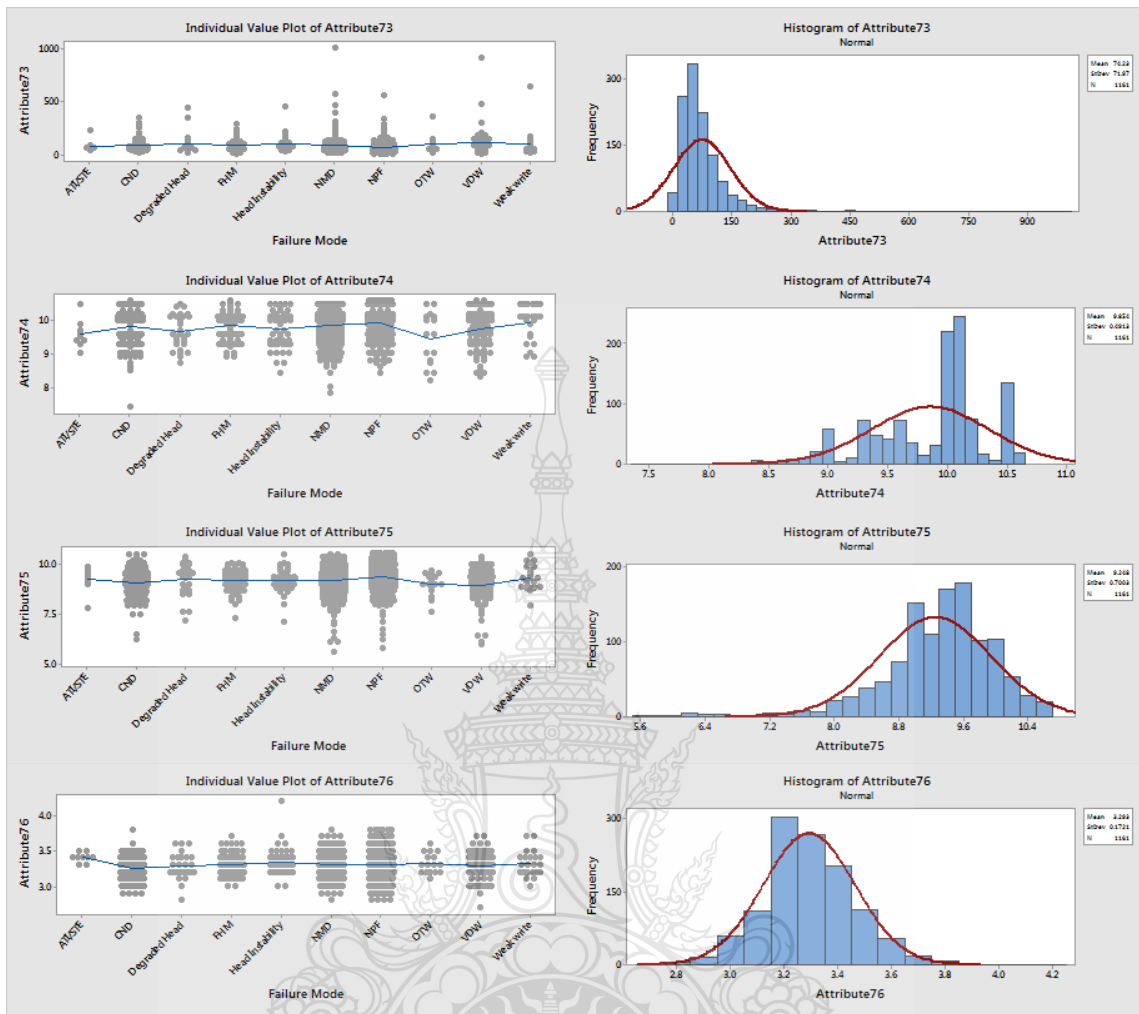






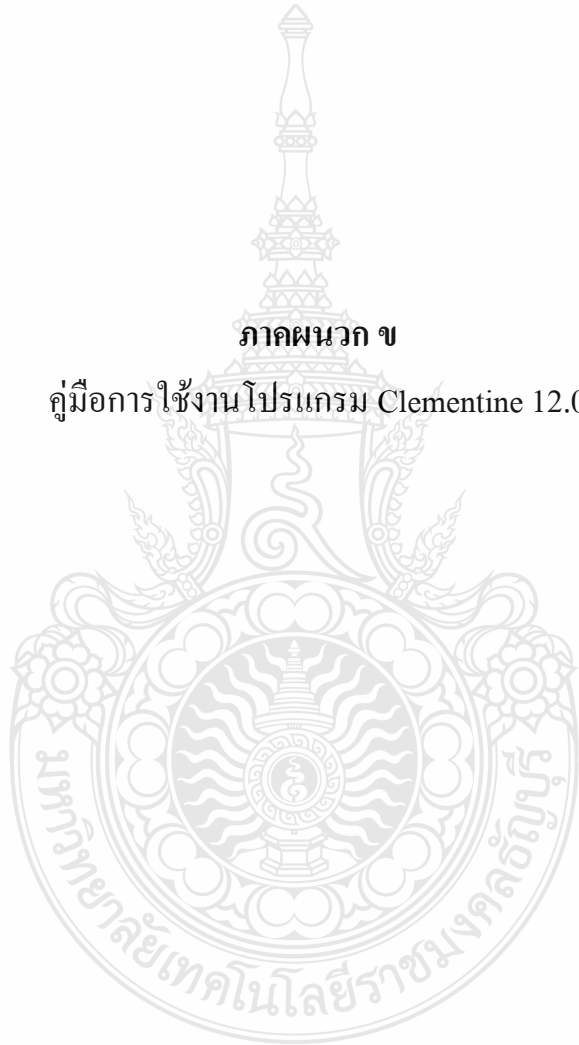






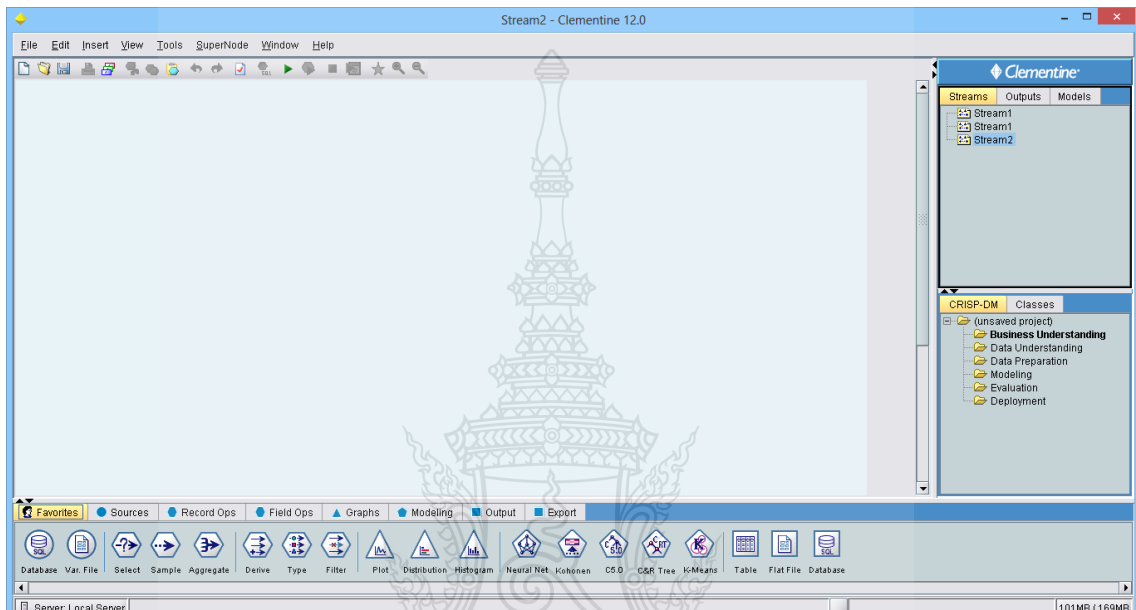
ภาคผนวก ข

คู่มือการใช้งานโปรแกรม Clementine 12.0



## โปรแกรม SPSS Clementine

โปรแกรม SPSS Clementine เป็นโปรแกรมที่ได้ถูกพัฒนาขึ้นโดยพนักงานของ Integral Solutions Ltd.(ISL) ซึ่ง Clementine นั้นเป็นผลิตภัณฑ์ที่รวบรวมเทคนิควิธีการต่าง ๆ ในการทำ Data Mining ให้เลือกใช้ตามความเหมาะสมอย่างมากมาย เช่น Neural Networks, Rule Induction, Kohonen Networks, Association Rule และ C5.0 เป็นต้น



รูปที่ ข.1 แสดงหน้าหลักของโปรแกรม Clementine 12.0

ส่วนประกอบหลักของโปรแกรม Clementine 12.0 มีดังนี้

Managers เป็นส่วนที่แสดง Streams ที่ใช้งาน

Nodes Palette เป็นส่วนที่แสดง Nodes ประเภทต่าง ๆ ได้แก่

Sources เป็นกลุ่ม Nodes สำหรับอ่านข้อมูลจากแหล่งข้อมูลต่าง ๆ

Record Ops เป็นกลุ่ม Nodes ที่ใช้จัดการเก็บเรคคอร์ดของข้อมูล เช่น การจัดเรียง การรวบรวมข้อมูล เป็นต้น

Field Ops เป็นกลุ่ม Nodes ที่ใช้สำหรับจัดการเก็บ Field ข้อมูล เช่น การกำหนดชนิดข้อมูล (Type) การแสดงข้อมูล (filter) เป็นต้น

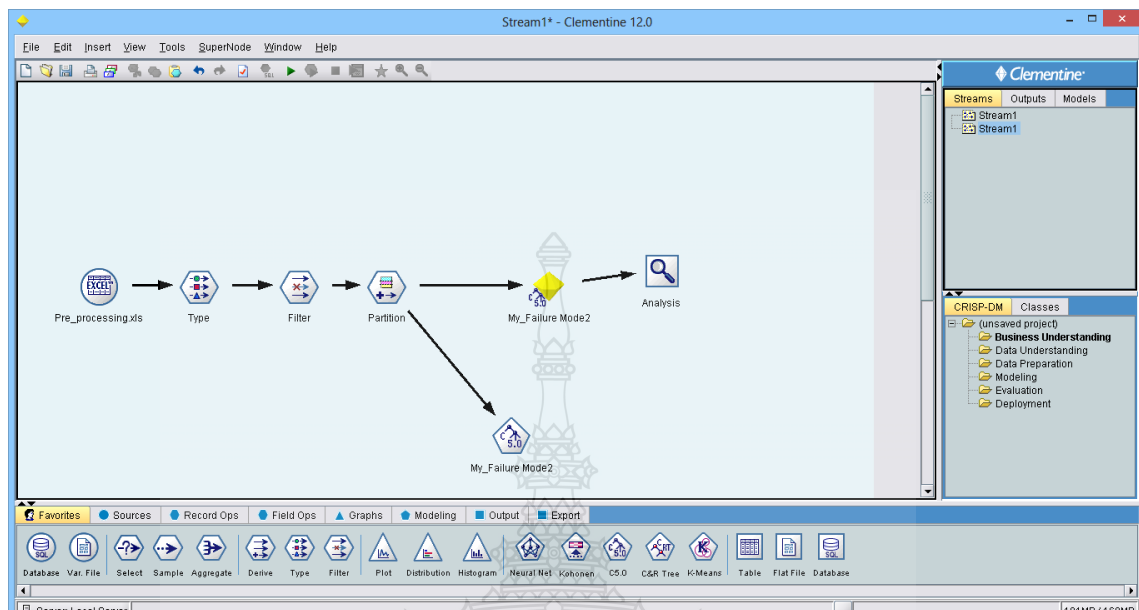
Graphs เป็นกลุ่ม Nodes ที่ใช้แสดงกราฟแบบต่าง ๆ

Modeling เป็นกลุ่ม Nodes สำหรับประมวลผลข้อมูลด้วยอัลกอริทึมแบบต่าง ๆ เช่น

Neural Networks C5.0 เป็นต้น

Output เป็นกลุ่ม Nodes ที่ใช้แสดงผลลัพธ์ในรูปแบบต่าง ๆ

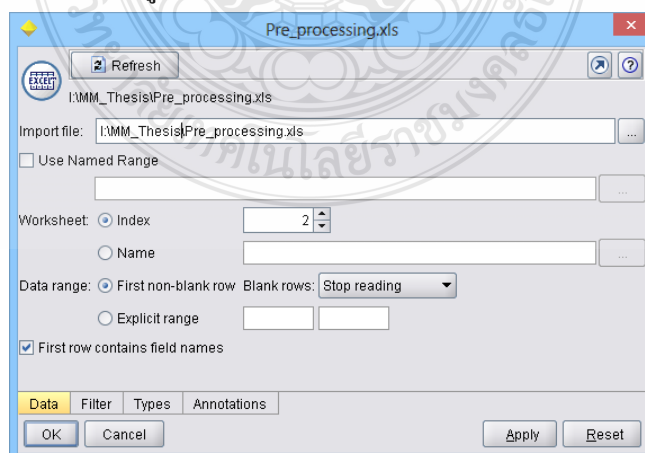
เมื่อจัดเตรียมข้อมูลเรียบร้อยแล้ว ขั้นตอนต่อมา คือ การพัฒนาตัวแบบโดยใช้โปรแกรม SPSS Clementine โดยเลือกใช้ อัลกอริทึม C5.0 ในการสร้างตัวแบบตามขั้นตอนดังต่อไปนี้



รูปที่ ข.2 การสร้างตัวแบบด้วย โปรแกรม Clementine 12.0

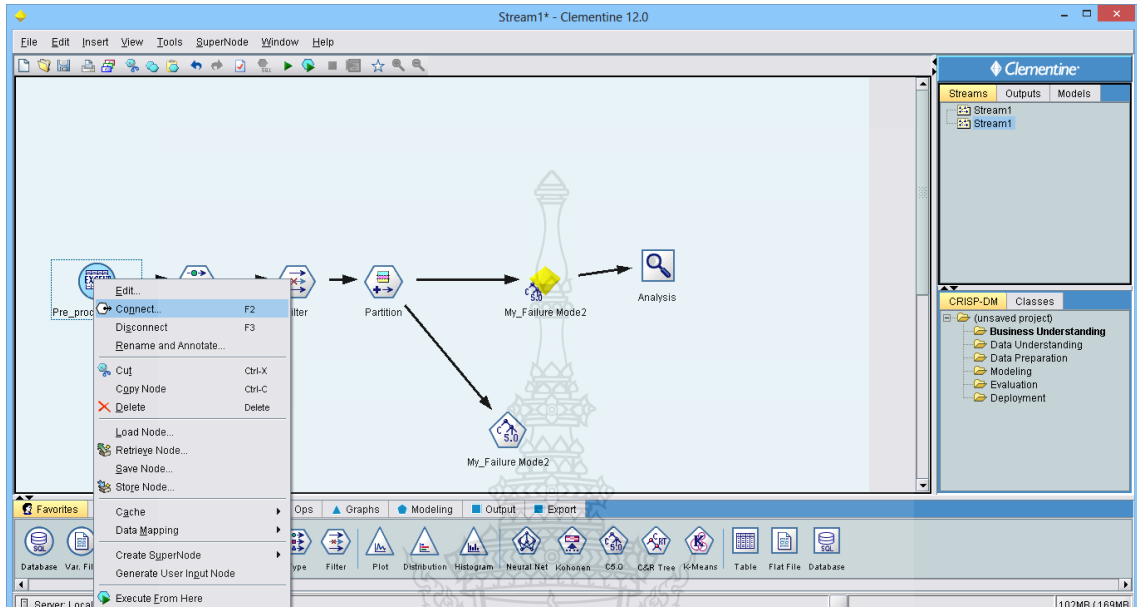
ขั้นตอนการสร้างตัวแบบด้วยอัลกอริทึม C5.0 ประกอบด้วยรายละเอียดดังนี้

1. การนำเข้าข้อมูล ใน โปรแกรม SPSS Clementine 12.0 สามารถนำข้อมูลเข้ามาประมวลผลจากแหล่งข้อมูลหลายรูปแบบด้วยกัน เช่น จากเท็กซ์ไฟล์ ฐานข้อมูล ไมโครซอฟท์เอกซ์เซล หรือ การเตรียมจากโปรแกรม SPSS ก็ได้ ในที่นี้เลือกวิธีการนำเข้าจากไมโครซอฟท์เอกซ์เซลดังแสดงในรูปที่ ข.3 โดยในขั้นตอนนี้สามารถกำหนดรายละเอียดต่าง ๆ เช่น ชนิดของข้อมูล (Types) ในเขตข้อมูลต่าง ๆ หรือการกรองข้อมูล (Filter)



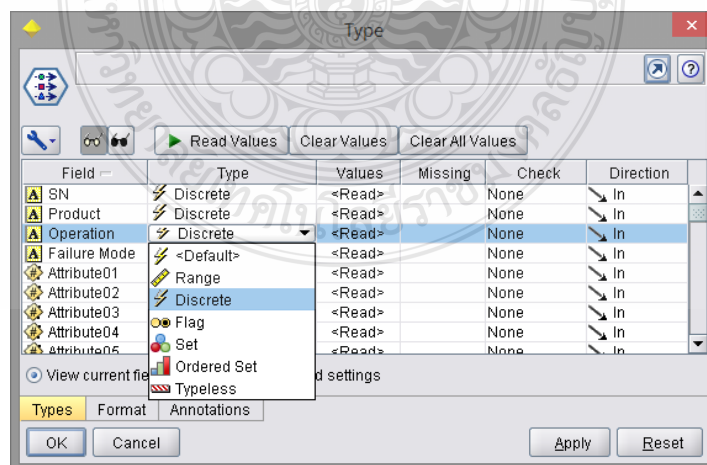
รูปที่ ข.3 การนำเข้าข้อมูลเข้าจากแหล่งข้อมูลประเภทไมโครซอฟท์เอกซ์เซล

2. การเชื่อมโยง Node เมื่อนำ Node ใด ๆ มาวางบนพื้นที่ทำการแล้ว ต้องทำการเชื่อมโยง Node ต่าง ๆ เข้าด้วยกันเพื่อให้สามารถประมวลผลได้ ดังรูปที่ ข.4 เมื่อ Node ต่าง ๆ เชื่อมโยงกันแล้ว จึงสามารถดำเนินการต่อไปได้



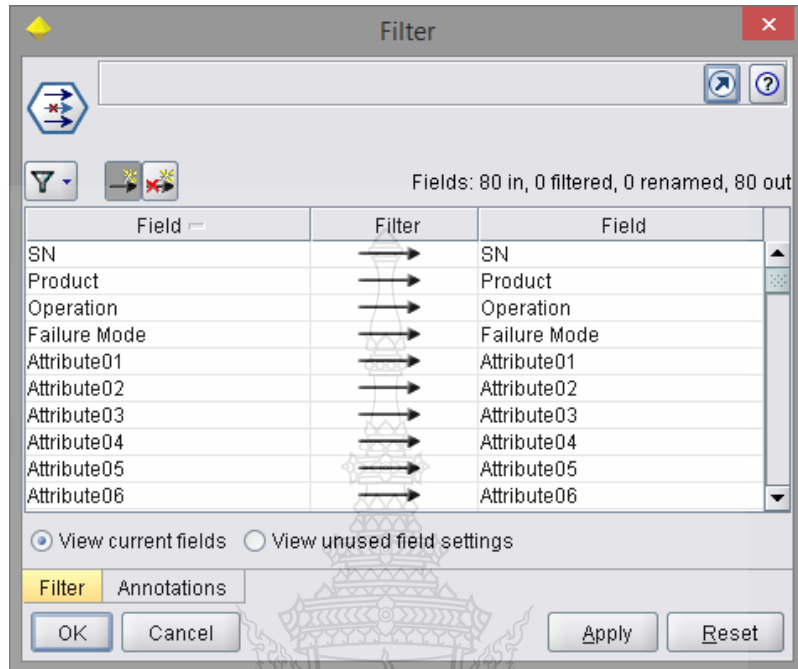
รูปที่ ข.4 การเชื่อมโยง Node

3. การกำหนดชนิดข้อมูล หลังจากเชื่อมโยง Node ข้อมูลเข้ากับ Type Node แล้วก็สามารถทำการกำหนดชนิดข้อมูลให้เหมาะสมได้ดังรูปที่ ข.5 โดยในส่วนของ Direction ได้กำหนดให้เขตข้อมูล Failure\_Mode มีค่าเป็น Out เพื่อให้ทำหน้าที่เป็น Class Label ส่วนเขตข้อมูลอื่นกำหนดให้เป็น In เพื่อให้ทำหน้าที่เป็น Input pattern



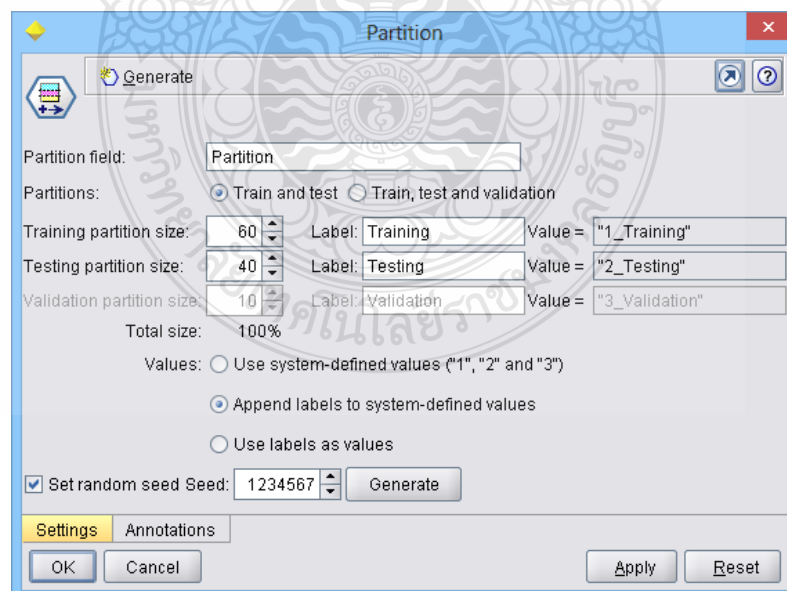
รูปที่ ข.5 การกำหนดชนิดของข้อมูล

4. การกรองข้อมูล หลังจากเชื่อมโยง Type Node ข้อมูลเข้ากับ Filter Node แล้วสามารถทำการกรองเพื่อเลือกเฉพาะเขตข้อมูลที่ต้องการใช้ในการประมวลผลต่อไป



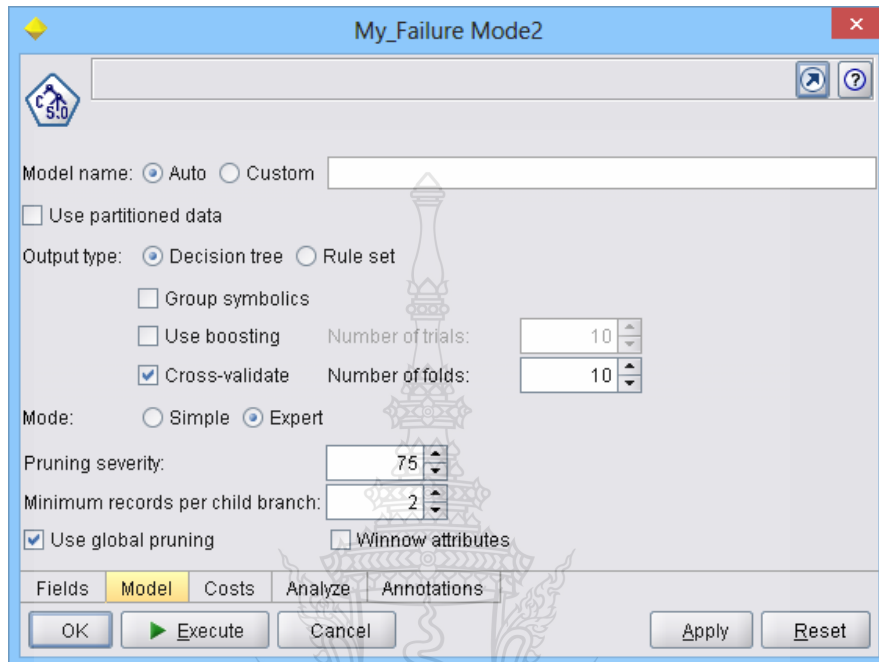
รูปที่ ข.6 การกรองของข้อมูล

5. การกำหนดตัวแบ่งข้อมูล โดยทำการแบ่งข้อมูลสำหรับชุดสอน (Training Partition) และชุดทดสอบ (Testing Partition)



รูปที่ ข.7 การกำหนดตัวแบ่งข้อมูล

6. การสร้างตัวแบบด้วยอัลกอริทึม C5.0 ในการสร้างตัวแบบด้วยอัลกอริทึม C5.0 นั้นมีขั้น  
คือ นำ C5.0 Node มาวางบนพื้นที่ทำการและเชื่อมโยงเข้ากับ Filter Node จากนั้นทำการกำหนด  
รายละเอียด ดังรูปที่ ข.8



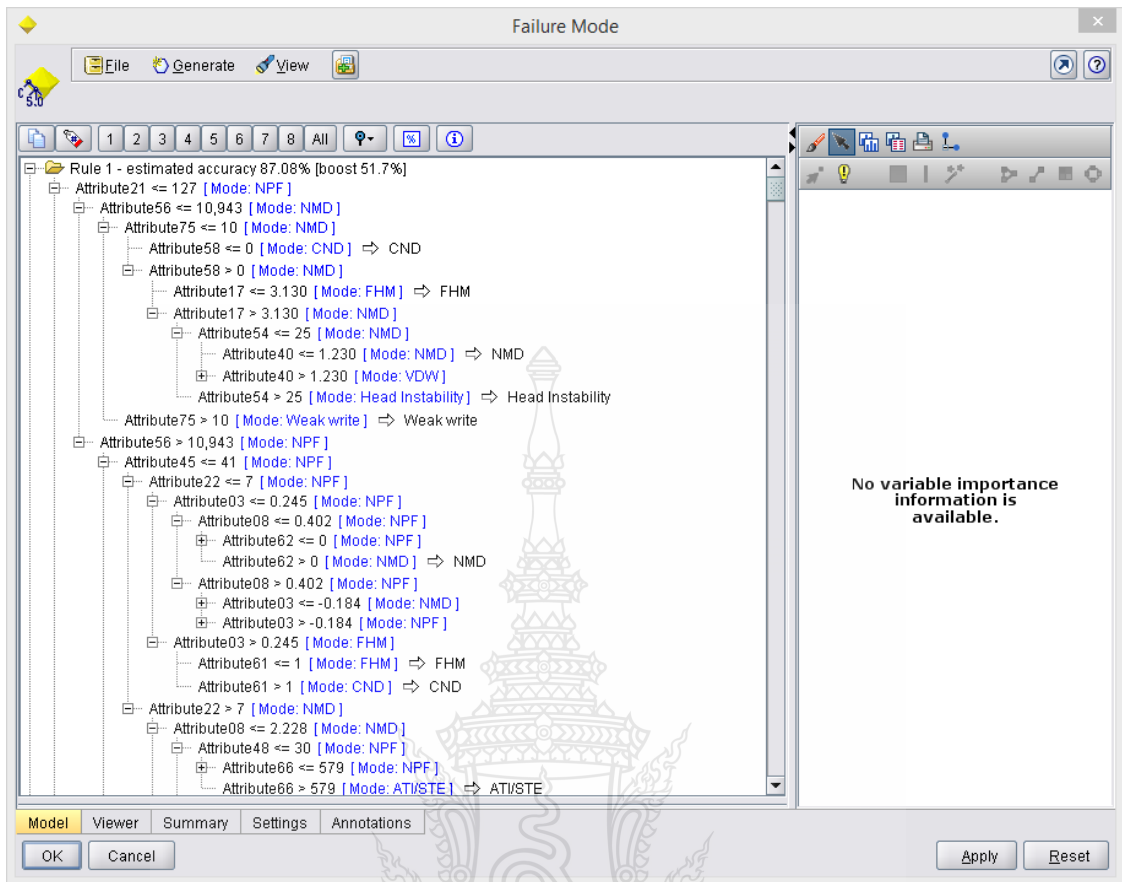
รูปที่ ข.8 การกำหนดรายละเอียดของ Node C5.0

ในการสร้างตัวแบบด้วยอัลกอริทึม C5.0 นั้น สามารถกำหนดรายละเอียดต่าง ๆ เพื่อให้ได้  
ตัวแบบที่มีประสิทธิภาพมากที่สุดก่อนที่จะนำไปใช้งานต่อไป ซึ่งเราสามารถเลือกให้ผลลัพธ์ออกมา  
เป็นต้นไม้ตัดสินใจ (Decision tree) หรือเป็นกฎการจำแนก (Rule set) ก็ได้ โดยการกำหนดที่ส่วน  
Output type สำหรับการทดลองนี้ได้เลือกผลลัพธ์เป็น Decision tree เพื่อนำตัวแบบที่ได้ไปทดสอบหา  
ประสิทธิภาพก่อนแล้วจึงสร้างกฎการจำแนกภายหลังจากที่ได้ตัวแบบที่น่าพอใจแล้ว

อีกส่วนหนึ่งที่มีความสำคัญมากเช่นกันคือ Mode ในการสร้างตัวแบบซึ่งสามารถกำหนดได้  
2 แบบได้คือ ง่าย (Simple) และขั้นสูง (Expert) สำหรับ Mode Expert นั้นเราสามารถกำหนด  
รายละเอียดเพิ่มเติมได้คือ เปอร์เซนต์ในการทำ Pruning และกำหนดจำนวนต่ำเรคคอร์ดต่ำสุดที่ยอม  
ให้มีการสร้างกฎได้ (Minimum records per child branch) ซึ่งสองส่วนนี้เป็นส่วนที่มีความสำคัญมาก  
ในการสร้างตัวแบบให้มีประสิทธิภาพตามที่ต้องการ

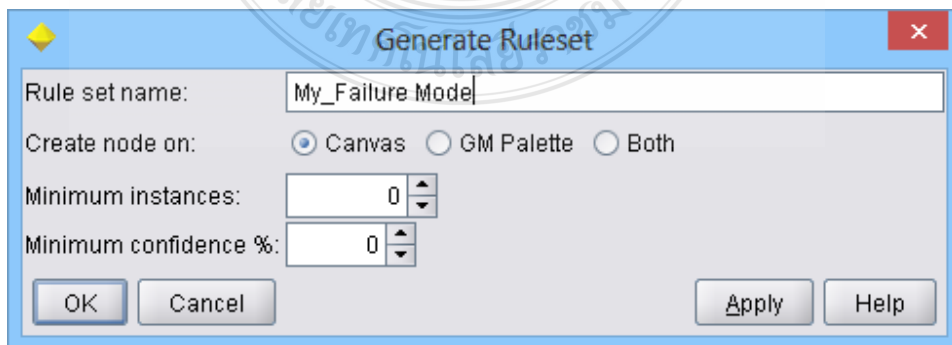
เมื่อทำการกำหนดรายละเอียดต่าง ๆ เรียบร้อยแล้วจึงสั่ง Execute จะได้ผลลัพธ์เป็นตัวแบบ  
ดังรูปที่ ข.9



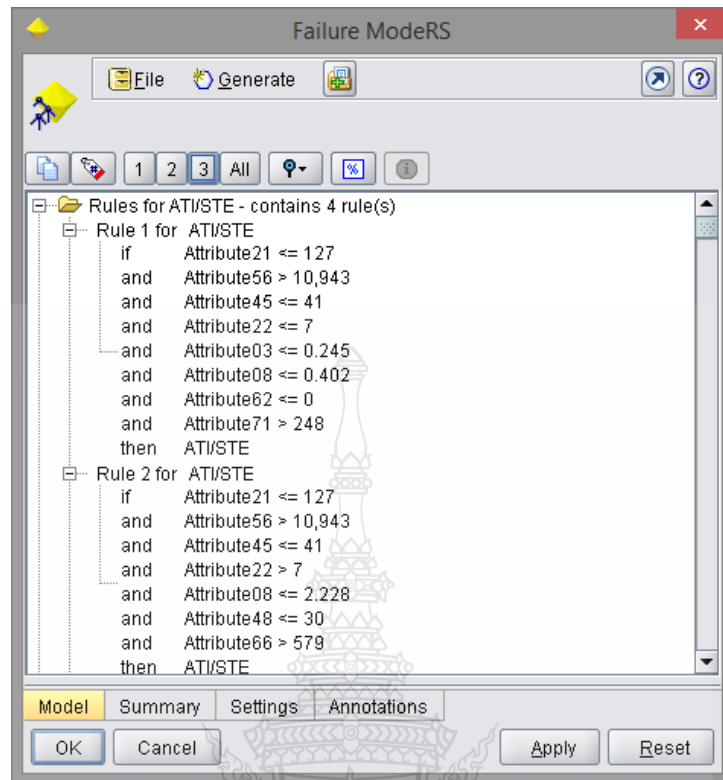


รูปที่ ข.9 ตัวแบบ (Model) ที่สร้างจากอัลกอริทึม C5.0

จากตัวแบบที่ได้จะแสดงอยู่ในรูปของโครงสร้างต้นไม้ (Tree) เราสามารถปรับการแสดงผลของต้นไม้ให้แสดงได้หลายระดับด้วยกัน จากนั้นถ้าต้องการสร้างกฎการจำแนกให้เลือกคำสั่ง Generate แล้วเลือกคำสั่ง Rule set จะปรากฏกรอบตอบโต้ดังรูปที่ ข.10 ให้ทำการกำหนดค่า Minimum instances และค่า Minimum confidence แล้วคลิกปุ่ม OK เพื่อสร้างกฎการจำแนกดังรูปที่ ข.11



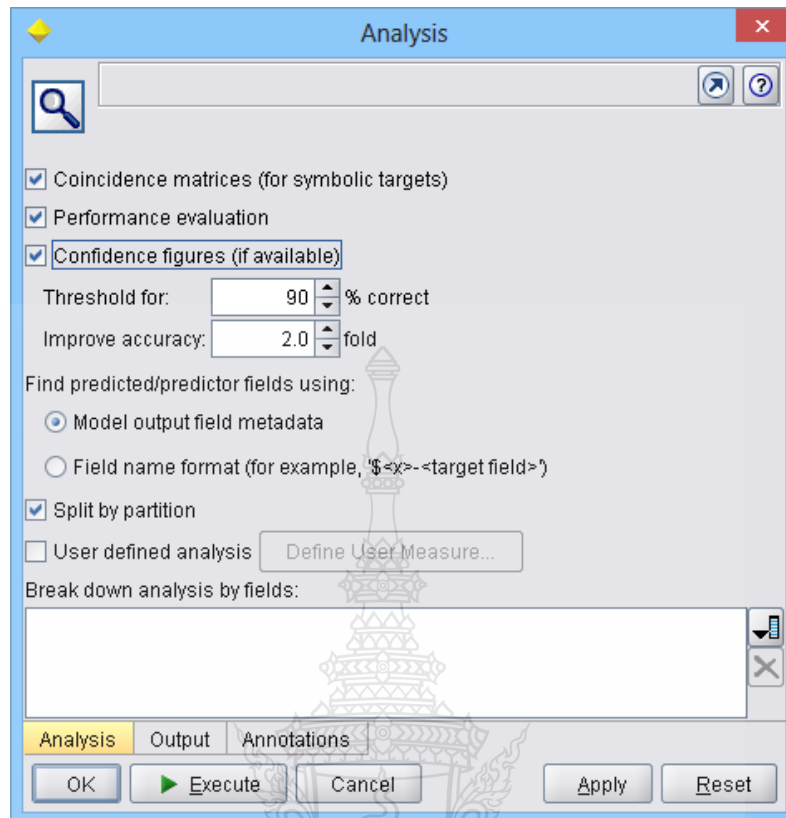
รูปที่ ข.10 การแปลงเป็นกฎการจำแนก



รูปที่ ข.11 กฎการจำแนก

การทดสอบความถูกต้องในการทำนายของตัวแบบ

1. นำข้อมูลที่เตรียมไว้สำหรับทดสอบตัวแบบเข้าสู่โปรแกรม โดยมีขั้นตอนเช่นเดียวกับขั้นตอนในการสร้างตัวแบบ
2. ทำการกำหนดชนิดข้อมูลและกรองข้อมูล โดยมีขั้นตอนเช่นเดียวกับขั้นตอนในการสร้างตัวแบบ
3. นำตัวแบบที่ต้องการทดสอบมาวางบนพื้นที่ทำการ จากนั้นทำการเชื่อมต่อ Node ต่าง ๆ เข้าด้วยกัน
4. นำ Node Analysis และ Node Matrix จากกลุ่ม Output มาวางบนพื้นที่และมีการกำหนดรายละเอียดต่าง ๆ ได้แก่ส่วนของ Analysis ดังรูปที่ ข.12 และผลลัพธ์ที่ได้ดังรูปที่ ข.13



รูปที่ ข.12 การกำหนดรายละเอียดของ Node Analysis

Results for output field Failure Mode

Comparing \$C-Failure Mode with Failure Mode

'Partition'	1_Training		2_Testing	
Correct	677	98.83%	475	99.79%
Wrong	8	1.17%	1	0.21%
Total	685		476	

Coincidence Matrix for \$C-Failure Mode (rows show actuals)

'Partition' = 1_Training	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	5	0	0	0	0	0	0	0	0	0
CND	0	74	0	0	0	0	0	0	0	0
Degraded Head	0	0	16	0	0	0	2	0	0	0
FHM	0	0	0	47	0	1	0	0	0	0
Head Instability	0	0	0	0	30	0	0	0	0	0
NMD	0	0	0	0	0	182	2	0	0	0
NPF	0	0	0	0	0	0	257	0	0	0
OTW	0	0	0	0	0	0	0	10	0	0
VDW	0	0	0	0	0	0	0	1	44	0
Weak write	0	0	0	0	0	0	0	2	0	12
'Partition' = 2_Testing	ATI/STE	CND	Degraded Head	FHM	Head Instability	NMD	NPF	OTW	VDW	Weak write
ATI/STE	4	0	0	0	0	0	0	0	0	0
CND	0	40	0	0	0	0	0	0	0	0
Degraded Head	0	0	9	0	0	0	1	0	0	0
FHM	0	0	0	19	0	0	0	0	0	0
Head Instability	0	0	0	0	16	0	0	0	0	0
NMD	0	0	0	0	0	146	0	0	0	0
NPF	0	0	0	0	0	0	182	0	0	0
OTW	0	0	0	0	0	0	0	6	0	0
VDW	0	0	0	0	0	0	0	0	45	0
Weak write	0	0	0	0	0	0	0	0	0	8

รูปที่ ข.13 การแสดงผลลัพธ์ของ Node Analysis



**ภาคผนวก ค**

**ผลงานตีพิมพ์เผยแพร่**

- ค.1** การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐานอุตสาหกรรมด้วยโครงข่ายประสาทเทียม,  
EENET 2014, 26-28 มีนาคม 2557, มารีไทม์ ปาร์คแอนสปารีสอร์ท, กระบี่
- ค.2** Hard Disk Drive Failure Mode Prediction based on Industrial Standard using Decision Tree  
Learning, ECTI-CON 2014, 14-17 May 2014, Nakhon Ratchasima, Thailand

ค.1 การตรวจสอบคุณภาพของฮาร์ดดิस्कภายใต้มาตรฐานอุตสาหกรรมด้วยโครงข่ายประสาทเทียม

**การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า**

**มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ ๖**

6<sup>th</sup> Electrical Engineering Network 2014  
of Rajamangala University of Technology Conference  
นวัตกรรมวิจัยแห่งเทคโนโลยี ตอบสนองเออีซีด้านโครงข่ายพลังงาน  
Technology Research Innovation for Responding  
to the Energy Network of AEC

**Volume II**

- อิเล็กทรอนิกส์ (EL)
- ไฟฟ้าสื่อสาร (CM)
- คอมพิวเตอร์และเทคโนโลยีสารสนเทศ (CP)
- ระบบควบคุมและการวัด (CT)
- การประมวลผลสัญญาณดิจิทัล (DS)
- งานวิจัยที่เกี่ยวข้องกับวิศวกรรมไฟฟ้า (GN)

๒๖ - ๒๘ มีนาคม พ.ศ. ๒๕๕๗  
ณ มาร์ทิม ปาร์คแอนสปาร์ตอร์ท จังหวัดกระบี่  
ดำเนินการโดย คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย

### รายนามผู้ทรงคุณวุฒิพิจารณาบทความ

รองศาสตราจารย์ ดร.มนตรี กาญจนะเคชะ	มหาวิทยาลัยสงขลานครินทร์
รองศาสตราจารย์ ดร.วิจิตร กิณเรศ	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
รองศาสตราจารย์ ดร.อนุวัฒน์ จางวนิชเลิศ	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
รองศาสตราจารย์ ดร.กิริติ ชยะกุลศิริ	มหาวิทยาลัยศรีปทุม
รองศาสตราจารย์ ดร.บัลลังก์ เนียมมณี	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
รองศาสตราจารย์ ดร.สันติ หวังนิพนานโต	สถาบันเทคโนโลยีปทุมวัน
รองศาสตราจารย์ ดร.เสถียร ธัญญศรีรัตน์	สถาบันเทคโนโลยีปทุมวัน
รองศาสตราจารย์ ดร.สมเกียรติ อุดมพรหมากุล	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
รองศาสตราจารย์ ดร.กานต์ เกิดชื่น	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
รองศาสตราจารย์ ดร.โกศล โอฬารไพโรจน์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
รองศาสตราจารย์ชัยณรงค์ วิเศษศักดิ์วิชัย	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
รองศาสตราจารย์สมพันธ์ อำพวัน	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าศรีราชา
ผู้ช่วยศาสตราจารย์ ดร.ณัฐชาจินดาเพชร	มหาวิทยาลัยสงขลานครินทร์
ผู้ช่วยศาสตราจารย์ ดร.ปัทมาภรณ์ อารีอุต	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
ผู้ช่วยศาสตราจารย์ ดร.ชัยวัฒน์ สากุล	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าศรีราชา
ผู้ช่วยศาสตราจารย์ ดร.กฤษณ์ชนม์ ภูมิภักดิ์พิชญ์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.สมชัย ทิริอุไรโรคม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.บุญยัง ปลั่งกลาง	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.จักรี ศรีนนท์ฉัตร	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.อำนาจ เรืองวารี	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.สำเร็จสินท่าไม้	มหาวิทยาลัยศรีปทุม
ผู้ช่วยศาสตราจารย์ ดร.จักรพงษ์ งามมิตร	สถาบันเทคโนโลยีปทุมวัน
ผู้ช่วยศาสตราจารย์ ดร.วิชัย ครอบกิจศิริ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
ผู้ช่วยศาสตราจารย์ ดร.ประมุข อุณหเลขกะ	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์ ดร.ปรีชา สาคะรังค์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์ ดร.ภักดิ์วัฒน์ จันทร์ศรี	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์ ดร.กฤตวิทย์ บัวใหญ่	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าอีสาน
ผู้ช่วยศาสตราจารย์ ดร.เจนศักดิ์ เอกบูรณะวัฒน์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้ารัตนโกสินทร์
ผู้ช่วยศาสตราจารย์ ดร.กันต์พงษ์ ศรีสถิตย์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้ารัตนโกสินทร์
ผู้ช่วยศาสตราจารย์ ดร.อรุณพล ป้อมสถิตย์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้ารัตนโกสินทร์
ผู้ช่วยศาสตราจารย์ ดร.ชาญชัย เศรษฐธรรมรงค์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.อุเทน คำน่าน	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.วันไชย คำแสน	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.พานิช อินต๊ะ	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.ประชา ยินยงกุล	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.วิมล ทองเล็ก	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ผู้ช่วยศาสตราจารย์ ดร.นาคยา ค้ายเรือง	มหาวิทยาลัยเกษตรศาสตร์
ผู้ช่วยศาสตราจารย์ ดร.บุญยฤทธิ์ ประสาทแก้ว	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.อาทิตย์ โสทร โยม	มหาวิทยาลัยสยาม

### รายนามผู้ทรงคุณวุฒิพิจารณาบทความ

ผู้ช่วยศาสตราจารย์ ดร.ยุทธชัย ศิลปวิจารณ์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ ดร.กฤษดา ชัยขันธ์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ศิวินท์ คั่นจันทู	มหาวิทยาลัยสงขลานครินทร์
ผู้ช่วยศาสตราจารย์ธวัชชัย ทางรัตน์สุวรรณ	มหาวิทยาลัยสงขลานครินทร์
ผู้ช่วยศาสตราจารย์ศิริชัย แลงเอม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์ทง ลานธารทอง	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
ผู้ช่วยศาสตราจารย์พิชญ์ คาราทพงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
ผู้ช่วยศาสตราจารย์สุกัญญา เนตรโพธิ์แก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
ผู้ช่วยศาสตราจารย์สรารุช ทองกุลภัทร์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์เฉลิมพล เรืองพัฒนาวิวัฒน์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์วารุณี กรุกพันธ์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
ผู้ช่วยศาสตราจารย์พันธ์พงศ์ อภิชาติกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
ผู้ช่วยศาสตราจารย์วิเศษ ทัพย์ประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผู้ช่วยศาสตราจารย์วิสุทธิ์ พงษ์พฤษภาค	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
ผู้ช่วยศาสตราจารย์สมคิด ลีลาชนะชัยพงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ ดร.วฤทธิ์ วิชากุล	มหาวิทยาลัยสงขลานครินทร์
อาจารย์ ดร.ณัฐภัทร พันธุ์คง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.ฉัตรชัย สุขพิทักษ์สกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.พฤษยน นินทนาวงศา	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.ศิริชัย เจริญล้ำเลิศ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.ภิตติวัฒน์ นิ่มเกิดผล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.วิเชียร อุแปแก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.สุริยา แก้วอาษา	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.กฤษดา สมสัย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.พรเทพ ปัญญาแก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.พิชัย อยู่เปล่า	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.นิธิโรจน์ พรสุวรรณเจริญ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.มธาท ทัตคร	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.นำพน พิพัฒน์ไพบูลย์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.สันทวัฒน์ ทองแดง	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.รสกสรร พลสุวรรณ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
อาจารย์ ดร.ณัฐโชติ รักไทยเจริญชีพ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.ณัฐพงศ์ พันธุ์นะ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.พนา คูสิตากร	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.สุรเชษฐ์ เคชพูน	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.พสิษฐ์ สุวรรณกิจการ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.ชลกาญจน์ วงศ์ก่อทรัพย์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ ดร.อุมารินทร์ แสงพานิช	มหาวิทยาลัยเกษตรศาสตร์
อาจารย์ ดร.ชาญณรงค์ บาลมงคล	จุฬาลงกรณ์มหาวิทยาลัย

### รายนามผู้ทรงคุณวุฒิพิจารณาบทความ

อาจารย์ ดร.ศรีสุตา ไชยทองสุข	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์ ดร.สมพร ศรีวัฒนพล	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์ ดร.ประจวบ อินระวงศ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์ ดร.พินิจ ศรีธร	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์ ดร.วรรณรีย์ วงศ์ไครรัตน์	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์ ดร.มงคล คูพิมาย	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์ ดร.ประสพโชค ให้ออกคำ	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ ดร.มนตรี สมดุลกนก	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ ดร.ไกรฤกษ์ เชยชื่น	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ ดร.พิชิต กิตติสุวรรณ	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ ดร.วิวัฒน์ ทิพพร	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.วิจักษ์ พรหมมี	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.นพพร พัวร์ ประกิติ	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.ชาอุยยุทธ กาญจนพิบูลย์	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.จักรกฤษณ์ เคลือบวัง	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.พลกฤษณ์ ทุนคำ	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.ยุพดี หัตถดิน	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.ขวัญชัย เอื้อวิริยานุกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.สุบงกช ไตโพนุลย์	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.อนนท์ นำอิน	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.สุภกิต แก้วดวงตา	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.จัดจตุรย์ ทองปรอน	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.ธีรศักดิ์ สมศักดิ์	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
อาจารย์ ดร.วุฒิวัฒน์ คงรัตน์ประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์ ดร.สายชล ชูดเจือจิน	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์ ดร.นิธิตวัฒน์ ชูสกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ ดร.พิทักษ์ บุญนุ่น	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ ดร.ปฏิมากร จันทร์พร้อม	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ ดร.ชยานิมฐ์ บุญสนิท	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
ดร.สุวิทย์ อังริยะเมต	สถาบันวิจัยวิทยาศาสตร์และเทคโนโลยีแห่งประเทศไทย
ดร.อรรด พยอมหอม	กรมไฟฟ้าพลังทดวง
อาจารย์พินิจ จิตจริง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์นิติพงศ์ ปานกลาง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์ภัทรพงศ์ อัญชันภาติ	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก
อาจารย์สมยศ สันติมาลัย	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก
อาจารย์วุฒิไกร จันทร์ขามเวียง	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก
อาจารย์ตะวัน ขุนอาสา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์มนัส บุญเทียรทอง	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์สาคร วุฒิพัฒน์พันธ์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร



### รายนามผู้ทรงคุณวุฒิพิจารณาบทความ

อาจารย์พูนศรี วรรณการ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์วาทรินทร์ ธัญสิประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์สุรสิทธิ์ ประกอบกิจ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์จตุรงค์ จตุรเชิดชัยสกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์อัครศักดิ์ วิริยกรรม	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ธนรัตน์ ตันมณีประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์อานนท์ สิงห์เสถียร	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์อภิษฐา ทองรักษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์ณิลนิต นิลาศ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อาจารย์พัชรนันท์ ศรีธนาอุทัยกร	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์มาลียา คังจิตเจมญา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์ธานี สมวงศ์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์ปฎิวัติ บุญมา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์อานนท์ พ่วงชิงงาม	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์วิสูตร คันทิรุ่งเรือง	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
อาจารย์ชูศักดิ์ กมลขันดิสร	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์บุญช่วย เจริญผล	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์เอกพล อนุสุเรนทร์	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์วินัย เมธาวิทิต	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์ชาญฤทธิ์ ชราสันติสุข	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
อาจารย์มงคล คำน่ำรุ่งครระกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์อัครศักดิ์ แจ็งสารกิจ	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์โสภา แซ่เฮ้ง	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ศิพล ฉ่ำเฉียวกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์กาญจน์ญา พานิชเจริญ	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ปกรณ์ ฐิติไพเราะ	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์ชาญยุทธ อุบายโกศล	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์พีรวัจน์ มีสุข	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
อาจารย์อภาพล มหาวีระ	มหาวิทยาลัยเทคโนโลยีราชมงคลอิสาน
อาจารย์นิกร แสงงาม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
อาจารย์คนโท ปานทองคำ	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์อัมพร บุญราม	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์พิทักษ์ สติสุวรรณะ	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์กิติ มินา	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ชัช สัตยรักษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ธีรพงษ์ ฉิมเพชร	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์ภราดร เรืองกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย
อาจารย์สำเอน ละอองโชค	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย

บทความสาขาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ (CP)		หน้า
<b>CP-C</b>		
CP22	การใช้งาน โปรแกรมประยุกต์กับเอ็นเอฟซีบนสมาร์ตโฟนเพื่อแจ้งเตือนการรับประทานยา สมโภชน์ กุศลธารารมย์ พงษ์พิสิฐ วุฒิมิขุ ไซติ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ	797
<b>CP-D</b>		
CP23	การศึกษาเปรียบเทียบความเสียหายทางการเงินของคณะวิศวกรรมศาสตร์ของมหาวิทยาลัยของรัฐบาล กับมหาวิทยาลัยในกำกับของรัฐละมหาวิทยาลัยเอกชน ในประเทศไทย สุภา ภูษา มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ	801
CP24	การจำลองการสร้างความถี่สูงโดยโพรงเส้นพ้องแบบวงแหวนขนาดเล็กด้วยโปรแกรม FDTD พงษ์ภู ไท อุดมอริยทรัพย์ มหาวิทยาลัยราชภัฏชัยภูมิ	805
CP25	การพัฒนาวิธีการจัดอันดับข้อมูลบนโครงข่ายแบบเพียร์ทูเพียร์ ชานนทร์ คงศรี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ	809
CP26	การตรวจสอบคุณภาพของสารเคลือบที่ได้มาตรฐานอุตสาหกรรมด้วยโครงข่ายประสาทเทียม ธนดล สุชาติพิงศ์ กฤษณ์ชนม์ ภูมิภักดิ์พิชญ์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี	813
<b>บทความสาขาระบบควบคุมและการวัด (CT)</b>		
<b>CT-A</b>		
CT01	การควบคุมอัตราส่วนข้อ/ขยายแบบสองทิศทางโดยใช้ตัวสังเกตการณ์สัญญาณรบกวน จากตัวกรองคาลมานสำหรับอุปกรณ์กวาดแพทย์ ดีเด่น เชาวลิศ มิตร์สันดีสุข มหาวิทยาลัยเกษตรศาสตร์	817
CT02	การเปรียบเทียบคุณภาพกำลังไฟฟ้าของอินเวอร์เตอร์จ่ายแรงดัน สำหรับขับเคลื่อนมอเตอร์ไฟฟ้า เหนี่ยวนำ 3 เฟส <sup>1</sup> คณพล คำปัญญา <sup>2</sup> บุญเรือง วงศ์ลาวัตร <sup>3</sup> เสถียร ชัยอนุสรณ์ <sup>1</sup> มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ <sup>2</sup> สถาบันเทคโนโลยีปทุมวัน	821
CT03	การวัดความต้านทานสูงโดยวิธีการวัดแบบโวลท์-แอมป์และสัมประสิทธิ์แรงเคลื่อน ชัยวัฒน์ เจษฎาจินต์ เนตรนพิศ คู่่มทุกทิส สถาบันนครวิทยานแห่งชาติ	825

## การตรวจสอบคุณภาพของฮาร์ดดิสก์ภายใต้มาตรฐานอุตสาหกรรมด้วยโครงข่ายประสาทเทียม Hard Disk Drive Quality Checking based on Industrial Standard using Artificial Neural Network

ธนดล สุชาติพงษ์ และ กฤษณ์ชนม์ ภูมิภักดิพิชญ์

สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

39 หมู่ 1 ถนนรังสิต-นครนายก อำเภอธัญบุรี จังหวัดปทุมธานี 12110 โทร. 0-2549-3571 โทรสาร 0-2549-3422 E-mail: Krischonme.b@en.rmutt.ac.th

### บทคัดย่อ

ฮาร์ดดิสก์เป็นอุปกรณ์ที่ใช้เก็บข้อมูล ในปัจจุบันความต้องการในการบันทึกข้อมูลมีความหลากหลายและไฟล์ข้อมูลมีขนาดใหญ่ ฮาร์ดดิสก์จึงจำเป็นต้องมีขนาดเพิ่มมากขึ้น ซึ่งในขั้นตอนการผลิตจะมีความซับซ้อน และมีการบันทึกข้อมูลทดสอบมากขึ้น ในกระบวนการผลิตจึงอาจมีของเสียเกิดขึ้น และอาจหลุดไปถึงมือผู้ใช้งานได้ ดังนั้นจึงมีการสุ่มเลือกฮาร์ดดิสก์จากการผลิตแต่ละครั้งมาทดสอบการจำลองการใช้งานในสภาพความเป็นจริง (Reliability Test) ซึ่งใช้เวลานานขึ้นอยู่กับความจุของฮาร์ดดิสก์ และจำนวนของฮาร์ดดิสก์ที่รอทำการทดสอบ หากมีการผลิตฮาร์ดดิสก์จำนวนมาก อาจทำให้ไม่สามารถทดสอบฮาร์ดดิสก์ได้ในเวลาที่กำหนด ซึ่งอาจทำให้ฮาร์ดดิสก์ส่งถึงมือผู้ใช้ช้ากว่ากำหนด และไม่สามารถเริ่มกระบวนการผลิตฮาร์ดดิสก์ในชุดถัดไปได้ ดังนั้นบทความนี้ได้นำเสนอวิธีการนำข้อมูลการทดสอบที่มีการบันทึกไว้ในกระบวนการผลิตของฮาร์ดดิสก์แต่ละตัวมาจัดกลุ่ม และปรับแต่งข้อมูล ใช้เทคนิค Pruning ในการปรับค่าโครงข่ายประสาทเทียมแบบ Back propagation ในการตรวจสอบคุณภาพการผลิต เพื่อลดขั้นตอน Reliability Test ซึ่งเป็นการเพิ่มความเร็วในการวิเคราะห์เพื่อแยกของเสียออกก่อนส่งมอบให้ลูกค้า ซึ่งผลที่ได้จากการทดลองโดยใช้การเรียนรู้แบบโครงข่ายประสาทเทียมแบบ Back propagation Algorithm จำนวน 18 ชั้นซ่อน มาตรวจสอบข้อมูลจาก Process Test 76 ค่า ได้ค่าความถูกต้องถึง 91.23%

คำสำคัญ: โครงข่ายประสาทเทียม, คุณภาพการผลิต, การคัดแยกข้อมูล

### Abstract

Hard Disk Drive (HDD) is a device used to store data. In the present there are high demand for storage diverse data. So there are essential and expansion increases. The production is complex process and there recorded a large amount of results in each production process. Which the production process, there are some failed drives and passed to customer. So, in each lots it random select HDD to simulate actual events. It takes a long time and the devices are not enough. Therefore, this paper presents the procedure to get data of each HDD to integration

and normalization using Pruning technique to optimize the neural network to help predict the production quality. Purpose to reduce the process of simulation actual use testing, increases the speed and efficiency of the analysis process before delivery to the customer. The results of the experiments using artificial neural network Back propagation Algorithm can be predicted optimal at number of hidden layer is 18 layers with accuracy 91.23% .

Keywords: Artificial Neural Networks, Quality Checking, Pattern Classification

### 1. บทนำ

ความก้าวหน้าทางด้านเทคโนโลยีสารสนเทศทำให้เกิดความต้องการฮาร์ดดิสก์เพื่อจัดเก็บข้อมูลเป็นจำนวนมาก โดยในปัจจุบันข้อมูลที่ถูกจัดเก็บในฮาร์ดดิสก์มีความสำคัญมาก เช่น ข้อมูลทางการค้าของบริษัทห้างร้านต่างๆ ข้อมูลการเงินส่วนบุคคล และข้อมูลทางราชการเป็นต้น ทำให้ผู้ใช้งานคาดหวังฮาร์ดดิสก์ที่มีคุณภาพสูงจากเหตุผลดังกล่าวทำให้อุตสาหกรรมการผลิตฮาร์ดดิสก์ต้องมีการพัฒนาเพิ่มขึ้นทั้งในด้านกำลังการผลิตและคุณภาพของสินค้า ซึ่งขั้นตอนการผลิตฮาร์ดดิสก์ในปัจจุบันจะมีการทดสอบคุณภาพในหัวข้อต่างๆ ตามมาตรฐานการทดสอบ (Process Test) [1-2] และหลังจากผลิตครบ Lot แล้วก็จะถูกสุ่มมาทดสอบคุณภาพอีกครั้งหนึ่งซึ่งเรียกว่า Reliability Test ก่อนส่งมอบให้ลูกค้าต่อไปซึ่งการทดสอบดังกล่าวมีความสำคัญมากเนื่องจากยังพบว่าฮาร์ดดิสก์ที่ไม่ผ่านการทดสอบ ถึงแม้ฮาร์ดดิสก์ดังกล่าวจะผ่าน In Process Test มาแล้วซึ่งหมายความว่าถ้าขาดการทำ Reliability Test อาจจะทำให้ฮาร์ดดิสก์ที่ไม่มีคุณภาพถูกส่งมอบให้ลูกค้า แต่การทำ Reliability Test ใช้เวลานาน เครื่องทดสอบดังกล่าวก็มีราคาสูง และที่สำคัญในแต่ละผลิตภัณฑ์สามารถสุ่มมาทดสอบได้เพียง 10-15% และยังตรวจจับฮาร์ดดิสก์ที่มีโอกาสเสียได้อีก 0.73% จากกลุ่มที่เข้าทดสอบ [3] ใน Reliability Test ในบทความนี้จึงนำเสนอการประยุกต์ใช้โครงข่ายประสาทเทียมในการทำหน้าที่ Reliability Test โดยให้โครงข่ายประสาทเทียมหาความสัมพันธ์ระหว่างค่าที่ได้จาก Process Test และผลการทำ Reliability Test และจากความสัมพันธ์ดังกล่าวทำให้สามารถ นำได้จาก

**บทความวิจัย**

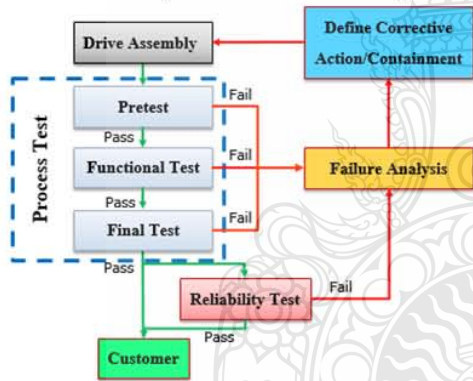
การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้ามหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 6

*Proceedings of the 6<sup>th</sup> Conference of Electrical Engineering Network of Rajamangala University of Technology 2014 (EENET 2014)*

Process Test ฮาร์ดดิสก์ใดๆ มาป้อนเข้าโครงข่ายประสาทเทียมและพยากรณ์ว่าฮาร์ดดิสก์ตัวดังกล่าวจะผ่าน Reliability Test หรือไม่ ซึ่งช่วยลดเวลาและค่าใช้จ่ายในการทำ Reliability Test และการใช้โครงข่ายประสาทเทียมดังกล่าวอาจกล่าวได้ว่า สามารถนำฮาร์ดดิสก์ทุกตัวผ่าน Reliability Test โดยไม่จำเป็นต้องสุ่มทดสอบ ซึ่งช่วยให้มั่นใจได้ว่าฮาร์ดดิสก์ที่ไม่ได้คุณภาพจะไม่ถูกส่งมอบให้ลูกค้า

**2. การตรวจสอบคุณภาพฮาร์ดดิสก์**

ในขั้นตอนการผลิตฮาร์ดดิสก์ประกอบด้วยกระบวนการต่างๆ ดังแสดงในรูปที่ 1 คือ Drives Assembly จะเป็นขั้นตอนการประกอบชิ้นส่วนของฮาร์ดดิสก์ ส่วน Pretest, Functional test และ Final test จะรวมอยู่ในกระบวนการทดสอบคุณภาพของกระบวนการผลิต และจะมีการสุ่มทดสอบที่ Reliability test อีกจำนวนหนึ่งก่อนส่งมอบให้ลูกค้า ซึ่งหากไม่ผ่านการทดสอบจะถูกส่งไปยัง Failure Analysis เพื่อวิเคราะห์อย่างละเอียดเพื่อหาสาเหตุของอาการเสียและแก้ไขได้อย่างถูกต้อง



รูปที่ 1 กระบวนการทดสอบฮาร์ดดิสก์ที่ใช้ในปัจจุบัน

**2.1 มาตรฐานการทดสอบในอุตสาหกรรมฮาร์ดดิสก์ (Process Test)**

ในขั้นตอน Process Test จะทำการทดสอบฮาร์ดดิสก์ และปรับแต่งค่าพารามิเตอร์ต่างๆ ให้สามารถใช้งานได้เหมาะสม โดยมาตรฐานที่สำคัญได้มีการกำหนดไว้ดังนี้ Start Motor ทดสอบช่วงเวลาในการตั้งแต่เปิดจนฮาร์ดดิสก์พร้อมใช้งาน Change motor speed ทดสอบความเร็วของมอเตอร์ Servo Seek ทดสอบระยะเวลาในการเข้าถึงข้อมูล Side Erasure Test ทดสอบการเขียนข้อมูล เพื่อป้องกันผลกระทบไปยังแทร็คข้างเคียง Thermal Asperity Scan ตรวจสอบหาตำแหน่งที่ผิดปกติบนแผ่นจานแม่เหล็ก ที่มีลักษณะเป็นรอยขีด ลึก หรือบุ๋มขึ้น Servo Flaw Scan ค้นหาข้อบกพร่องของสัญญาณเซอร์โว Contact Detect วัดระยะห่างระหว่างหัวอ่านและจานแม่เหล็ก Head Stability ตรวจสอบ

ความไม่เสถียรของหัวอ่าน โดยการอ่านเขียนข้อมูลเป็นเวลานานๆ หลายๆ ครั้ง Resonance ตรวจสอบความถี่เรโซแนนซ์และขนาดของสัญญาณบนหัวอ่าน Skip Write Detect เป็นการตรวจจับตำแหน่งที่เขียนไม่สมบูรณ์ Bit Error Rate by Zone ทดสอบจำนวนบิตที่เกิดการส่งผิดพลาดเทียบกับจำนวนที่ส่งทั้งหมดในแต่ละโซนข้อมูล Hot Cold Write Test ทดสอบการอ่าน/เขียนข้อมูลในอุณหภูมิสูงและต่ำ Read Primary List เป็นการตรวจสอบตำแหน่งที่ไม่สามารถอ่านเขียนข้อมูลได้ เพื่อตัดทิ้ง SMART Attribute Check ตรวจสอบค่าแอตทริบิวต์สะสมต่างๆ ที่อยู่ใน SMART ให้อยู่ในสเป็ค เช่น จำนวนครั้งการใช้งาน อุณหภูมิ ค่ากระแสที่ใช้ควบคุมหัวอ่าน ค่าความเสถียรของสัญญาณ จำนวนบิตที่เกิดการส่งผิดพลาด จำนวนการเขียนข้อมูลที่เกิดพลาด จำนวนการอ่านข้อมูล จำนวนการเขียนข้อมูล Zone Based Transfer Rate Test ทดสอบอัตราการส่งผ่านข้อมูล Head Resistance Test ทดสอบค่าความต้านทานของหัวอ่านข้อมูล

**2.2 มาตรฐานการสุ่มทดสอบคุณภาพ (Reliability Test)**

ในขั้นตอน Reliability Test [4] จะทำการทดสอบฮาร์ดดิสก์ในสภาวะการใช้งานต่างๆ เพื่อให้มั่นใจได้ว่าเมื่อลูกค้านำไปใช้งานแล้วจะไม่เกิดปัญหาขึ้น ในแต่ละการทดสอบจะมีการเก็บข้อมูลการทดสอบรวมถึงค่าพารามิเตอร์ต่างๆ ของฮาร์ดดิสก์ ซึ่งอาจเป็นข้อมูลในการระบุปัญหาที่จะเกิดขึ้นในอนาคตได้ โดยใช้การเปรียบเทียบกับค่าที่กำหนดไว้ (Threshold) จากทีมพัฒนา (Design center) ในบางครั้งข้อมูลการทดสอบของการทดสอบเดียวอาจไม่เพียงพอในการระบุปัญหาของฮาร์ดดิสก์จึงจำเป็นต้องใช้ข้อมูลจากการทดสอบอื่นๆ ร่วมด้วย เช่น การเปิด/ปิดฮาร์ดดิสก์หลายๆ ครั้งติดต่อกัน การเปิดทิ้งไว้เป็นเวลานานๆ การทดสอบอ่าน/เขียนข้อมูลในอุณหภูมิที่แตกต่างกัน หรือการทดสอบจ่ายกระแสไฟสูงสุด/ต่ำสุด ตามสเป็คที่กำหนดไว้ โดยมีลำดับการทดสอบดังตารางที่ 1

ตารางที่ 1 ลำดับขั้นตอนการสุ่มทดสอบคุณภาพ

Test Sequence	
1. Drive Power On	10. Write Pattern
2. Write Read Perform	11. Read Pattern
3. System File Load	12. Random Write
4. Voltage High Low	13. RAM Miss-compare
5. Temperature High Low	14. Write Zero Pattern
6. Short Drive Self-Test	15. Read Zero Verify
7. Low Duty Cycle	16. Long Drive Self-Test
8. OS Write Test	17. SMART Reset
9. OS Read Test	18. Drive Power Off

## บทความวิจัย

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้ามหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 6

Proceedings of the 6<sup>th</sup> Conference of Electrical Engineering Network of Rajamangala University of Technology 2014 (EENET 2014)

### 2.3 ประเภทการเสียของฮาร์ดดิสก์ (Failure Mode)

ผลจากการทดสอบจาก Reliability Test ได้มีการกำหนดไว้เป็น Failure Mode [5] ที่สำคัญดังนี้

- NMD (New Media Defect) คือ เกิดจากหัวอ่านไปสัมผัสกับแผ่นบันทึกข้อมูล ทำให้ไม่สามารถบันทึกข้อมูลบริเวณดังกล่าวได้
- FHM (Fly Height Modulation) คือ อาการเสียลักษณะนี้เกิดจากการแกว่งของหัวอ่าน ทำให้อ่านข้อมูลบนแผ่นจานแม่เหล็กได้ไม่บ่อยดี
- Weak Write คือ เกิดจากสิ่งปนเปื้อน บดบังขณะที่หัวเขียนกำลังเขียนข้อมูลลงบนแผ่นจานแม่เหล็ก ทำให้ข้อมูลบริเวณนั้นขาดหายๆ ไป
- Degraded Head คือ หัวอ่าน/เขียน มีปัญหา ทำให้ไม่สามารถอ่านข้อมูลจากแผ่นจานแม่เหล็กได้
- OTF (Off-track Write) คือ เกิดจากความผิดพลาดในการบันทึกข้อมูลมีแรงกระทำภายนอกหรือการกำหนดจากอุปกรณ์ภายใน
- ATI คือ แทร็คข้างเคียง โคนสนามแม่เหล็กจากแทร็คที่กำลังเขียนรบกวนในขณะที่เขียนข้อมูลลงบนแผ่นจานแม่เหล็ก

### 2.4 ข้อมูลสมาร์ท (S.M.A.R.T)

สมาร์ท หรือ S.M.A.R.T (Self-Monitoring Analysis and Reporting Technology) [6-7] เป็นกระบวนการทำงานที่บันทึกค่าต่างๆ ของฮาร์ดดิสก์โดยใช้เฟิร์มแวร์ในไบออส มีวิธีการทำงานโดยการจับเก็บข้อมูลที่เกิดขึ้นตั้งแต่วันที่ผลิต อุณหภูมิ จำนวนความผิดพลาดต่างๆ ที่พบภายในฮาร์ดดิสก์ในระหว่างที่ Self-Calibration ซึ่งจะถูกบันทึกเป็นระยะๆ ในขณะที่ใช้งาน ทั้งที่แก้ไขได้และไม่ได้ เพราะในระหว่างใช้งานอาจมีการกระทบกระเทือนหรืออุณหภูมิสูงหรือต่ำเกินไป

### 2.5 การวิเคราะห์ข้อมูล

การนำข้อมูลที่เก็บรวบรวมได้ จากกลุ่มตัวอย่างหรือจากประชากรการวิจัยจำนวนหนึ่ง มาจำแนกเพื่อหาค่าสถิติที่เป็นตัวแทนของข้อมูลทั้งหมด เรียกว่า การวัดแนวโน้มเข้าสู่ส่วนกลาง ค่าสถิติที่นิยมกัน ได้แก่ ค่าเฉลี่ย (Mean)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

การวัดการกระจาย การวัดแนวโน้มเข้าสู่ส่วนกลางบอกได้เพียงค่าที่เป็นตัวแทนของข้อมูลชุดนั้นๆ แต่ไม่สามารถบอกได้ว่าข้อมูลเหล่านั้นมีค่าใกล้เคียงกัน หรือแตกต่างกันมากน้อยเพียงใด โดยสามารถวัดได้ด้วยตัวสถิติหลายตัว [8] ซึ่งสามารถคำนวณหาได้ดังสมการด้านล่างตามลำดับ

ค่าพิสัย (Range)

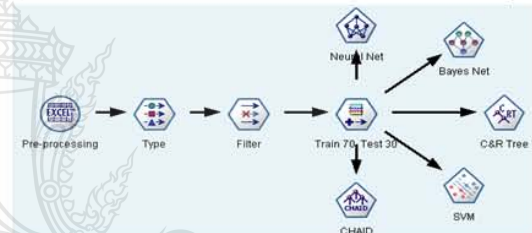
$$R = X_{\max} - X_{\min} \quad (2)$$

ค่าความเบี่ยงเบนมาตรฐาน (Standard deviation)

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (3)$$

### 2.6 การเลือกใช้อัลกอริทึม

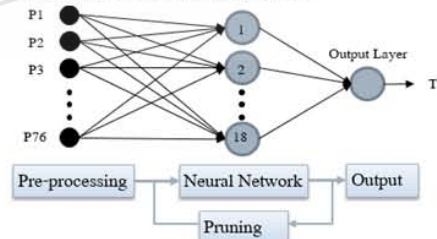
ข้อมูลที่ผ่านกระบวนการปรับแต่งข้อมูลให้มีมาตรฐานเดียวกันแล้วแบ่งข้อมูลเป็นสองชุดสำหรับการเรียนรู้ 70% และการทดสอบ 30% เพื่อหาอัลกอริทึมที่เหมาะสมที่สุด โดยเลือกอัลกอริทึมที่นำมาใช้ทดสอบดังนี้ Neural Network, Bayesian Network, Cart Tree, Support Vector Machine (SVM) และ Chaid Tree ในทดลองนี้ผู้วิจัยได้ทำการปรับแต่งพารามิเตอร์ต่างๆ เพื่อให้ได้ผลลัพธ์ที่ดีที่สุดของแต่ละอัลกอริทึม เนื่องจากข้อมูลที่ใช้ในการทดลองอยู่ในรูปแบบชุดข้อมูลต่อเนื่อง (Continuous) ซึ่งเหมาะสมกับการใช้ Neural Network มากที่สุด



รูปที่ 2 โครงสร้างและอัลกอริทึมในการทดสอบ

### 3. การนำโครงข่ายประสาทเทียมมาประยุกต์ใช้

บทความนี้ได้สร้างโครงข่ายประสาทเทียม โดยมี input เป็นข้อมูลที่ได้จาก Process Test จำนวน 76 ค่า โดยผ่านการปรับข้อมูลที่ได้จากการกำหนดค่าตัวแปรอินพุต (Pre-processing) ให้มีมาตรฐานเดียวกัน (Normalization) ด้วยวิธีหาค่าเฉลี่ย ค่าพิสัย ค่าความเบี่ยงเบนมาตรฐาน และมีเอทาคูต 1 ค่าที่จะแสดงผลเป็น 0 หรือ 1 แทนการ FAIL และ PASS ของการทำ Reliability Test ใช้ชั้นซ่อนจำนวน 18 ชั้นซ่อน Tan-Sigmoid เป็น Transfer Function ของแต่ละชั้นซ่อนและใช้ Pruning Method [9] โดยเริ่มจากการใช้โครงข่ายประสาทเทียมขนาดใหญ่และลดขนาดของส่วนที่ไม่จำเป็นลง ชั้น Output ดังแสดงในรูปที่ 3



รูปที่ 3 สถาปัตยกรรมโครงข่ายประสาทเทียมและเทคนิค Pruning

**บทความวิจัย**

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้ามหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 6

*Proceedings of the 6<sup>th</sup> Conference of Electrical Engineering Network of Rajamangala University of Technology 2014 (EENET 2014)*

จากนั้นได้เตรียมข้อมูลของฮาร์ดดิสก์เพื่อฝึกสอนโครงข่ายประสาทเทียมเรียนรู้จำนวน 586 ตัว ซึ่งประกอบด้วยตัวที่ "PASS" 315 ตัว และตัวที่ "FAIL" 271 ตัว ในขั้นตอน Reliability Test การฝึกสอนใช้ Back propagation Algorithm ทำการฝึกสอนให้โครงข่ายประสาทเทียมปรับค่า Weight และ Bias โดยอัตโนมัติโดยใช้โปรแกรมสำเร็จรูป Clementine12 จนได้ค่าเหมาะสมที่สามารถจัดรูปแบบความสัมพันธ์ของค่าใน Process Test และ ผลของ Reliability Test ได้อย่างถูกต้อง

**4. ผลการทดลอง**

ผลการทดลองที่ได้สร้างโครงข่ายประสาทเทียมที่สามารถจัดรูปแบบความสัมพันธ์ของ ค่าใน Process Test และ ผลของ Reliability Test ได้อย่างถูกต้องแล้ว จะต้องทดสอบการทำงานด้วยการป้อนข้อมูล Process Test ของฮาร์ดดิสก์ที่ไม่เคยนำมาฝึกสอนเพื่อให้โครงข่ายประสาทเทียมพยากรณ์ผลการทดสอบ Reliability Test และตรวจสอบว่าตรงกับผลการทดสอบด้วยเครื่องทดสอบหรือไม่อย่างไร ในบทความนี้ได้ใช้ข้อมูลของฮาร์ดดิสก์ที่เตรียมไว้สำหรับทดสอบ 228 ตัว ซึ่งประกอบด้วยตัวที่ "PASS" 124 ตัว และตัวที่ "FAIL" 104 ตัว มาทดสอบการทำงานของโครงข่ายประสาทเทียมพบว่า สามารถพยากรณ์ผลการทดสอบ Reliability Test ได้โดยมีความผิดพลาด 8.77% ซึ่งผลที่ได้จากการทดลองได้ใช้อัลกอริทึมอื่นมาเปรียบเทียบกับตารางที่ 2

ตารางที่ 2 ผลลัพธ์ของแต่ละอัลกอริทึมจากข้อมูลชุดเดียวกัน

Algorithm	Training			Testing		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
SVM	506	80	86.35%	195	33	85.53%
Bayesian	525	61	89.59%	201	27	88.16%
Neural Net	528	58	90.10%	208	20	91.23%
CHAID	522	64	89.08%	204	24	89.47%
C&R Tree	495	91	84.47%	195	33	85.53%

ตารางที่ 3 เปรียบเทียบผลลัพธ์ระหว่าง Reliability และ Neural Network

	Neural Network Prediction						
		Training			Testing		
		FAIL	PASS	Accuracy	FAIL	PASS	Accuracy
Reliability	FAIL	232	39	85.61%	92	12	88.46%
Result	PASS	19	296	93.96%	8	116	93.55%

**5. สรุป**

Reliability Test เป็นกระบวนการที่สำคัญในการควบคุมคุณภาพของฮาร์ดดิสก์ก่อนที่จะส่งมอบให้ลูกค้าซึ่งกระบวนการดังกล่าวต้องใช้เวลาและค่าใช้จ่ายสูงในการทดสอบ บทความนี้จึงได้นำเสนอการประยุกต์ใช้โครงข่ายประสาทเทียมในการช่วยพยากรณ์ผลของการทำ Reliability Test ของฮาร์ดดิสก์โดยใช้ข้อมูล Process Test เป็นข้อมูลป้อนเข้า โครงข่ายประสาทเทียมที่ถูกฝึกสอนแล้วจะสามารถลดค่าความสัมพันธ์ระหว่าง ค่าใน In Process Test และ ผลของการทดสอบ Reliability Test และจากการทดสอบพบว่าโครงข่ายประสาทเทียมสามารถพยากรณ์ ผลการทดสอบ Reliability Test ได้อย่างมีประสิทธิภาพโดยมีความผิดพลาด 8.77%

**เอกสารอ้างอิง**

- [1] ดนัยพงศ์ เชษฐโชติศักดิ์, "การพยากรณ์ผลสภาพการผลิตของกระบวนการผลิตฮาร์ดดิสก์ด้วยโครงข่ายประสาทเทียม", *KKU Engineering Journal* Vol. 34 No. 4 (465 - 475), 2007.
- [2] M.T. White, "Characterization and Compensation of Disturbances Caused by Head-Disk Contact in Hard Disk Drives" in *IEEE transactions on magnetic recording*, 2007.
- [3] Hard disk drive reliability and MTBF / AFR [Online]. Available: [http://knowledge.seagate.com/articles/en\\_US/FAQ/174791en?language=en\\_US](http://knowledge.seagate.com/articles/en_US/FAQ/174791en?language=en_US)
- [4] B. D. Strom, S. C. Lee, G. W. Tyndall and A. Khurshudov, "Hard disk drive reliability modeling and failure prediction," *IEEE transactions on magnetic recording*, 2006.
- [5] Ramangkul T, Ponsawad J, "Hard Disk Drive Failure Mode Prediction from SMART Attribute using Data Mining Method," *DST-CON 2011*; January 2011.
- [6] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado, "Improved disk drive failure warnings," in *IEEE transactions on reliability*, September 2002.
- [7] Dave Anderson, Hard Disk SMART Drives [Online]. Available: <http://smartlinux.sourceforge.net/smart/index.php>
- [8] ขดิศร แก้วกิติ, "การปรับปรุงเวลาในการทดสอบฮาร์ดดิสก์ การควบคุมกระบวนการด้วยวิธีการทางสถิติ และวิเคราะห์ความสามารถของกระบวนการ", *EECON-32*, 2552.
- [9] T. Waraporn, K. Yuttana, "Customer Failure Modes Prediction for Hard Disk Drives using Neural Networks Rank-Level Fusion", *ECTICON2011*, Khon Kaen, Thailand, p.476.

พ.2 Hard Disk Drive Failure Mode Prediction based on Industrial Standard using Decision Tree Learning



## Reviewer List

Full Name	Affiliation Name	Country
Kar Hang Leung	Universiti Tunku Abdul Rahman	Malaysia
Yi Zheng	Shanghai Jiaotong University	China
Ming Lee Gan	Universiti Tunku Abdul Rahman	Malaysia
Bok Min Goi	Universiti Tunku Abdul Rahman	Malaysia
Chaofang Hu	Tianjin University	China
Yadong Zhao	China Univesiy of Petroleum	China
Hua Yang	Ocean University of China	China
Pramin Artrit	Khon Kaen University	Thailand
Surachai Chaitusaney	Chulalongkorn	Thailand
Pitipong Chanloha	Chulalongkorn University	Thailand
Wudhichai Assawinchaichote	King Mongkut's University of Technology Thonburi	Thailand
Vithida Chongsuphajaisiddhi	King Mongkut's University of Technology Thonburi	Thailand
Tussanai Parthornratt	Assumption University	Thailand
Chirdpong Deelertpaiboon	King Mongkut's University of Technology North Bangkok	Thailand
Tanit Malakorn	Naresuan University	Thailand
Supawan Ponpitakchai	Naresuan University	Thailand
Wanchak Lenwari	King Mongkut University of Tecnology Thonburi	Thailand
Yuanyuan Zou	East China University of Science and Technology	China
Supatana Auethavekiat	Chulalongkorn University	Thailand
Parinya Sanguansat	Panyapiwat Institute of Management (PIM)	Thailand
Darun Kesrarat	Assumption University	Thailand
Poonlap Lamsrichan	Kasetsart University	Thailand
Narong Aphiratsakun	Assumption University	Thailand
Saowaluck Kaewkamnerd	National Electronics and Computer Technology Center (NECTEC)	Thailand
Sopon Phumeechanya	Silpakorn University	Thailand
Hock Guan Goh	Universiti Tunku Abdul Rahman	Malaysia
Mongkol Ekpanyapong	Asian Institute of Technology, Thailand	Thailand



Bundit Thipakorn	King Mongkut's University of Technology Thonburi	Thailand
Angkoon Phinyomark	University of Calgary	Canada
Nattee Niparnan	Chulalongkorn University	Thailand
Mutita Songjun	Naresuan University	Thailand
Jonglak Pahasa	University of Phayao	Thailand
Worawat Nakawiro	King Mongkut Institute of Tech, Ladkrabang	Thailand
Yaser Qudaih	Kyushu Institute of Technology	Japan
Nopporn Leeprechanon	Thammasat University	Thailand
Pichai Aree	Thammasat University	Thailand
Supakit Chotigo	King Mongkut's University of Technology Thonburi	Thailand
Sanchai Dechanupaprittha	Kasetsart University	Thailand
Komsan Hongesombut	Kasetsart University	Thailand
Sumate Naetiladdanon	King Mongkut's University of Technology Thonburi	Thailand
Pichaya Chaipanya	Srinakharinwirot University	Thailand
Chayanit Bunsanit	Rajamangala University of technology Srivijaya	Thailand
Monthippa Uthansakul	Suranaree University of Technology	Thailand
Tanapong Khomyat	Rajamangala University of Technology Lanna (RMUTL)	Thailand
Piyaporn Krachodnok	Suranaree University of Technology	Thailand
Wanwisa Thaiwirot	King Mongkut's University of Technology North Bangkok	Thailand
Chalernpol Chamsripinyo	National Electronics and Computer Technology Center (NECTEC)	Thailand
Phongsak Keeratiwintakorn	King Mongkut's University of Technology North Bangkok	Thailand
Peerapon Siripongwutikorn	King Mongkut's University of Technology Thonburi	Thailand
Nattapong Swangmuang	Chiang Mai University	Thailand
Boon Hee Soong	Nanyang Technological University	Singapore
Kazi Mohiuddin Ahmed	Asian Institute of Technology, Thailand	Thailand
Suwat Pattaramalai	King Mongkut's University of Technology Thonburi	Thailand
Pasd Putthapipat	Assumption University	Thailand
Wiroon Sriborriurux	Burapha University	Thailand
Patrachart Komolkiti	Assumption University	Thailand
Chitapong Wechtaisong	Rajabhat University (Nakhon Ratchasima)	Thailand
Peerapong Uthansakul	Suranaree University of Technology	Thailand
Tanapat Anusasamornkul	King Mongkut's University of Technology North Bangkok	Thailand

Phisan Kaewprapha	Thammasat University	Thailand
Apirath Limmanee	Burapha University	Thailand
Surachet Kanprachar	Naresuan University	Thailand
Kornkamol Thakulsukanant	Assumption University	Thailand
Nararat Ruangchajitupon	Khon Kaen University	Thailand
Thumrongrat Amornraksa	King Mongkut's University of Technology Thonburi	Thailand
Warangkhana Kimpan	King Mongkut's Institute of Technology Ladkrabang	Thailand
Pirawat Watanapongse	Kasetsart University	Thailand
Pongpisit Wuttidittachotti	King Mongkut's University of Technology North Bangkok	Thailand
Rattapoom Waranusast	Naresuan University	Thailand
Somchai Limsiroratana	Prince of Songkla University	Thailand
Chatchanan Jandaeng	Walailak University	Thailand
Ladda Preechaveerakul	Prince of Songkla University	Thailand
Boontawee Suntasirivaraporn	Sirindhorn International Institute of Technology	Thailand
Thanaset Thosdeekoraphat	Suranaree University of Technology	Thailand
Suchaisri Lion	University of Phayao	Thailand
Lunchakorn Wuttisittikulij	Chulalongkorn University	Thailand
Werapon Chiracharit	King Mongkut's University of Technology Thonburi	Thailand
Wattanapong Kurdthongmee	Walailak University	Thailand
Watcharin Kaewapichai	Prince of Songkla University	Thailand
Boonsong Sutapun	Suranaree University of Technology	Thailand
Nayot Kurukitkoson	Burapha University	Thailand
Pornchai Supnithi	King Mongkut Institute of Tech, Ladkrabang	Thailand
Prayoot Akkaraekthalin	King Mongkut's University of Technology North Bangkok	Thailand
Prapun Suksompong	Sirindhorn International Institute of Technology	Thailand
Pisit Vanichchanunt	King Mongkut's University of Technology North Bangkok	Thailand
Apichai Bhatranand	King Mongkut's University of Technology Thonburi	Thailand
Keattisak Sripimanwat	National Electronics and Computer Technology Center (NECTEC)	Thailand
Cholwich Nattee	Sirindhorn International Institute of Technology	Thailand
Peerapol Yuvapoositanon	Mahanakorn University of Technology	Thailand
Amorn Jirasereeamornkun	Mahanakorn University of Technology	Thailand
Theerayos Wiangtong	Mahanakorn University of Technology	Thailand

Nalin Sidahao	Mahanakorn University of Technology	Thailand
Jirayuth Mahattanakul	Mahanakorn University of Technology	Thailand
Apisak Worapishet	Mahanakorn University of Technology	Thailand
Suwit Kiravittaya	Naresuan University	Thailand
Ekapon Siwapornsathain	King Mongkut's University of Technology Thonburi	Thailand
Pinit Kumhom	King Mongkut's University of Technology Thonburi	Thailand
Nitikarn Nimsuk	Thammasat University	Thailand
Wijitra Petchakit	Walailak University	Thailand
Nattha Jindapetch	Prince of Songkla University	Thailand
Sirirat Petsangri	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Sanya Mitaim	Thammasat University	Thailand
Phanumas Khumsat	Prince of Songkla University	Thailand
Cheng Wai Kheng	Universiti Tunku Abdul Rahman	Malaysia
Yen Kaow Ng	Universiti Tunku Abdul Rahman	Malaysia
Vijit Kinnares	King Mongkut's Institute of Technology Ladkrabang	Thailand
Siroj Sirisukprasert	Kasetsart University	Thailand
Surin Khomfoi	King Mongkut's Institute of Technology Ladkrabang	Thailand
Niphath Jantharamin	Naresuan University	Thailand
Sakorn Po Ngam	King Mongkut's University of Technology Thonburi	Thailand
Yuttana Kumsuwan	Chiang Mai University	Thailand
Anuwat Jangwanitert	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Piyadanai Pachanapan	Naresuan University	Thailand
Suratsavadee Korkua	Walailak University	Thailand
Pisit Liutanakul	King Mongkut's University of Technology North Bangkok	Thailand
Warachart Saekok	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Sompob Polmai	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Uthen Kamnarn	Rajamangala University of Technology Chiangmai	Thailand
Chanin Bunraksananusorn	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Surapong Suwankawin	Chulalongkorn University	Thailand
Terdpun Choogon	Nakhon Sawan Rajabhat University	Thailand
Yong Kheng Goh	Universiti Tunku Abdul Rahman	Malaysia
Kamol Kaemarungsi	National Electronics and Computer Technology Center (NECTEC)	Thailand

Kriangkrai Maneerat	Suranaree University of Technology	Thailand
Chuwong Phongcharoenpanich	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Duangarhit Srimoon	Rangsit University	Thailand
Chokchai Sangdao	Faculty of Engineering, Mahanakorn University of Technology	Thailand
Kongpol Areerak	Suranaree University of Technology	Thailand
Virasit Imtawil	Khon Kaen University	Thailand
Kongpan Areerak	Suranaree University of Technology	Thailand
Wipawee Hattagam	Suranaree University of Technology	Thailand
Mingmanas Sivaraksa	Mahidol University	Thailand
Nittaya Kerdprasop	Suranaree University of Technology	Thailand
Lalita Narupiyakul	Mahidol University	Thailand
Srisupa Palakvangsa Na Ayudhya	Mahidol University	Thailand
Thanwadee Sunetnanta	Mahidol University	Thailand
Tomoaki Sato	Hirosaki University	Japan
Montri Phothisonothai	The University of Tokyo	Japan
Sorawat Chivapreecha	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Suree Funilkul	King Mongkut's University of Technology Thonburi	Thailand
Sirapat Boonkrong	King Mongkut's University of Technology North Bangkok	Thailand
Waraporn Jirapanthong	Dhurakij Pundit University	Thailand
Wichian Chutimaskul	King Mongkut's University of Technology Thonburi	Thailand
Rattana Wetprasit	Prince of Songkla University	Thailand
Sermasak Uatrongjit	Chiang Mai University	Thailand
Decha Wilairat	Mahidol University	Thailand
Warakorn Kasamesuwan	King Mongkut Institute of Tech, Ladkrabang	Thailand
Apinunt Thanachayanont	King Mongkut Institute of Tech, Ladkrabang	Thailand
Rardchawadee Silapunt	King Mongkut's University of Technology Thonburi	Thailand
Mongkol Konghirun	King Mongkut University of Tecnology Thonburi	Thailand
Nisai Fuengwarodsakul	King Mongkut's University of Technology North Bangkok	Thailand
Issarachai Ngamroo	King Mongkut Institute of Tech, Ladkrabang	Thailand
Viboon Chunkag	King Mongkut's University of Technology North Bangkok	Thailand
Sally E Goldin	King Mongkut's University of Technology Thonburi	Thailand
Tanasanee Phienthrakul	Mahidol University	Thailand

Supawat Supakwong	Thammasat University	Thailand
Piyasawat Navaratana Na Ayudhya	King Mongkut's University of Technology Thonburi	Thailand
Tianhong Pan	Jiangsu University	China
Ali Ajami	Azarbaijan Shahid Madani University	Azerbaijan
Cattareeya Suwanasri	King Mongkut's University of Technology North Bangkok	Thailand
Komson Daroj	Ubon Ratchathani University	Thailand
Vuttiapon Tarateeraseth	Srinakharinwirot University	Thailand
Kohji Higuchi	University of Electro-Communications	Japan
Thanaphong Thanasaksiri	Chiang Mai University	Thailand
Kulyos Audomvongseree	Chulalongkorn	Thailand
Chakphed Madtharad	Provincial Electricity Authority (PEA)	Thailand
Phumin Kirawanich	Mahidol University	Thailand
Naruemon Wattanapongsakorn	King Mongkut's University of Technology Thonburi	Thailand
Supakit Prueksaaroon	Thammasat University	Thailand
Natawut Nupairoj	Chulalongkorn University	Thailand
Chakarida Nukoolkit	King Mongkut's University of Technology Thonburi	Thailand
Vajirasak Vanijja	King Mongkut's University of Technology Thonburi	Thailand
Narongrit Waraporn	King Mongkut's University of Technology Thonburi	Thailand
Tota Mizuno	University of Electro-Communications	Japan
Hojo Masahide	Tokushima University	Japan
Nattavit Chayavanich	King Mongkut's University of Technology Thonburi	Thailand
Amnat Suksri	Khon Kaen University	Thailand
Tasaneey Chayavanich	King Mongkut's University of Technology Thonburi	Thailand
Somchat Jiriwibhakorn	KMITL	Thailand
Somyot Kaitwanidvilai	KMITL	Thailand
Sompob Polmai	KMITL	Thailand
Wirote Buaklee	Provincial Electricity Authority (PEA)	Thailand
Panachit Kittipanya ngam	TRUE CORPORATION PUBLIC COMPANY LIMITED	Thailand
Chaiyod Pirak	King Mongkut's University of Technology North Bangkok	Thailand
Wilaiporn Lee	King Mongkut's University of Technology North Bangkok	Thailand
Parichat Sernwuthisarn	Kasetsart University at Si Racha Campus	Thailand
Kunagorn Kunavut	Assumption University	Thailand

Pollakrit Toonkum	Rajamangala University of Technology Lanna (RMUTL)	Thailand
Vorapoj Patanavijit	Assumption University	Thailand
Santitham Prom on	King Mongkut's University of Technology Thonburi	Thailand
Yuttapong Jiraksopakun	King Mongkut's University of Technology Thonburi	Thailand
Raungrong Suleesathira	King Mongkut's University of Technology Thonburi	Thailand
Toshiaki Kondo	Sirindhorn International Institute of Technology	Thailand
Pitikhate Sooraksa	King Mongkut Institute of Tech, Ladkrabang	Thailand
Suree Pumrin	Chulalongkorn University	Thailand
Teerapat Sanguankotchakorn	Asian Institute of Technology, Thailand	Thailand
Teerakiat Kerdcharoen	Mahidol University	Thailand
Tuchsanai Ploysuwan	SIAM UNIVERSITY	Thailand
Vanvisa Chutcahvong	King Mongkut Institute of Tech, Ladkrabang	Thailand
Sartra Wongthanavasu	Khon Kaen University	Thailand
Prapong Prechaprarnwong	King Mongkut University of Tecnology Thonburi	Thailand
Shuyu Xiang	Xidian university	Thailand
Dahmmaet Bunnjaveht	Thammasat University	Thailand
Theekapun Charoenpong	Srinakharinwirot University	Thailand
Apiwat Lekuthai	Chulalongkorn University	Thailand
Charnchai Pluempitiwiriyaewj	Chulalongkorn University	Thailand
Nisachon Tangsangiumvisai	Chulalongkorn University	Thailand
Chalie Charoenlarnpparut	Sirindhorn International Institute of Technology	Thailand
Yongyut Chonbodeechalermroong	Naresuan University	Thailand
Prakamkiat Youngkong	FIBO	Thailand
Wiklom Teeraparbkajorndet	Prince of Songkla University	Thailand
Sarawuth Chaimool	King Mongkut's University of Technology North Bangkok	Thailand
Rangsan Wongsan	Suranaree University of Technology	Thailand
Steven Gordon	Sirindhorn International Institute of Technology	Thailand
Kittisak Kerdprasop	Suranaree University of Technology	Thailand
Chaodit Aswakul	Chulalongkorn University	Thailand
Aduwati Sali	Universities Putra Malaysia	Malaysia
Chutiporn Anutariya	Shinawatra University	Thailand
Dulpichet Rerkpreedapong	Kasetsart University	Thailand

Parnjit Damrongkulkamjorn	Kasetsart University	Thailand
Chawasak Rakpenthai	University of Phayao	Thailand
Arthait Sode Yome	SIAM UNIVERSITY	Thailand
Naebboon Hoonchareon	Chulalongkorn	Thailand
Chanarong Banmongkol	Chulalongkorn University	Thailand
Chowarit Mitsantisuk	Kasetsart University	Thailand
Siwapon Srisophon	Kasetsart University	Thailand
Prayoth Kumsawat	Suranaree University of Technology	Thailand
Chanchai Thongsopa	Suranaree University of Technology	Thailand
Poompat Saengudomlert	Asian Institute of Technology, Thailand	Thailand
Chiranut Sangiamsak	Khon Kaen University	Thailand
Apirat Siritaratiwat	Khon Kaen University	Thailand
Settawit Poochaya	Suranaree University of Technology	Thailand
Akaraphunt Vongkunghae	Naresuan University	Thailand
Wisam Patchoo	Bangkok University	Thailand
Suwatchai Kamonsantiroj	King Mongkut's University of Technology North Bangkok	Thailand
Machigar Ongtang	Dhurakij Pundit University	Thailand
Yuttasart Nitipaichit	Cat Telecom Public Company Limited	Thailand
Luepol Pipanmaekaporn	King Mongkut's University of Technology North Bangkok	Thailand
Benchaphon Limthanmaphon	King Mongkut's University of Technology North Bangkok	Thailand
Akara Prayote	King Mongkut's University of Technology North Bangkok	Thailand
Yaowadee Temtanapat	Thammasat University	Thailand
Gridaphai Sriharee	King Mongkut's University of Technology North Bangkok	Thailand
Worasit Choochaiwattana	Dhurakij Pundit University	Thailand
dana blouin	Sirindhorn International Institute of Technology	Thailand
Yoichi Nakaguro	Sirindhorn International Institute of Technology	Thailand
Sam Banani	Sirindhorn International Institute of Technology	Thailand
Karin Sumongkayothin	Sirindhorn International Institute of Technology	Thailand
Thiri The Wut Yee	Sirindhorn International Institute of Technology	Thailand
Panuwat Janpugdee	Chulalongkorn University	Thailand
Danai Torrungrueng	Asian University	Thailand
Denchai Worasawate	Dept. Electrical Eng. Kasetsart University	Thailand

Chatree Mahatthanajatuphat	King Mongkut's University of Technology North Bangkok	Thailand
Akkarat Boonpoonga	King Mongkut's University of Technology North Bangkok	Thailand
Yan Zhao	Chulalongkorn University	Thailand
Suramate Chalermwisutkul	King Mongkut's University of Technology North Bangkok	Thailand
Boonsri Kaewkham Ai	Chiang Mai University	Thailand
Kittisak Phaebua	King Mongkut's University of Technology North Bangkok	Thailand
Phongsatorn Chomtung	King Mongkut's University of Technology North Bangkok	Thailand
Itthisek Nilkhamhang	Sirindhorn International Institute of Technology	Thailand
Taworn Benjanarasuth	King Mongkut's Institute of Technology Ladkrabang	Thailand
Supachai Vorapojpisut	Thammasat University	Thailand
Diew Koolpiruck	King Mongkut's University of Technology Thonburi	Thailand
Paisarn Muneesawang	Naresuan University	Thailand
Panomkhawn Riyamongkol	Naresuan University	Thailand
Rangsarit Vanijjirattikhan	NECTEC	Thailand
Sudchai Boonto	KMUTT	Thailand
Jitkomut Songsiri	Chulalongkorn University	Thailand
Thananchai Leephakpreeda	Sirindhorn International Institute of Technology (SIIT), Thammasat University	Thailand
Manop Wongsaisuwan	Chulalongkorn University	Thailand
Mongkol Saejia	Prince of Songkla University	Thailand
Chonlameth Arpnikanondt	King Mongkut's University of Technology Thonburi	Thailand
Kalika Suksomboon	KDDI	Japan
Chaiyachet Saivichit	Chulalongkorn University	Thailand
Waleed Soliman Mohammed	Bangkok University	Thailand
Weerawat Khawsuk	Chulachomklao Royal Military Academy	Thailand
Karel Sterckx	Bangkok University	Thailand
Pornchai Mongkolnam	King Mongkut's University of Technology Thonburi	Thailand
Bunthit Watanapa	King Mongkut's University of Technology Thonburi	Thailand
Athikom Roeksabutr	Mahanakorn University of Technology	Thailand
Touchai Angchuan	Prince of Songkla University	Thailand
Nattaphat Promsuwanna	Seagate Technolog (Thailand)	Thailand
Chatchai Khunboa	Khon Kaen University	Thailand
Sanan Srakaew	King Mongkut's University of Technology Thonburi	Thailand



Punpiti Piamsanga	Kasetsart University	Thailand
Yoshihiro Matsui	Tokyo National College of Technology	Japan
Koonlachat Meesublak	NECTEC	Thailand
Chanchana Thanachayanont	MTEC	Thailand
Satidchoke Phosaard	Suranaree University of Technology	Thailand
Chanwit Kaewkasi	Suranaree University of Technology	Thailand
Kultida Rojviboonchai	Chulalongkorn University	Thailand
Krit Athikulwongse	National Electronics and Computer Technology Center (NECTEC)	Thailand
Hideki Ayano	Tokyo National College of Technology	Japan
Somsak Vanitanunchai	Suranaree University of Technology	Thailand
Thipwan Fangsuwannarak	Suranaree University of Technology	Thailand
Pornpimon Chayratsami	King Mongkut's Institute of Technology Ladkrabang (KMITL)	Thailand
Boonruang Marungsri	Suranaree University of Technology	Thailand
Chutima Prommak	Suranaree University of Technology	Thailand



<b>Process Control .....</b>	<b>149</b>
Paper ID: 1049 .....	150
<i>Robust stabilizing multi-period repetitive controllers for multiple-input/multiple-output time-delay plants</i>	
Tatsuya Sakanushi, Yun Zhao, Kou Yamada, Natnari Smithimedhin	
Paper ID: 1119 .....	151
<i>Disturbance Observer based Friction Compensator for a DC Motor</i>	
M. K. C. Dinesh Chinthaka, R. U. G. Punchihewa, A.M.Harsha S. Abeykoon	
Paper ID: 1173 .....	152
<i>A Model-Reference Sliding Mode for Dual-Stage Actuator Servo Control in HDD</i>	
Siwaphon Sonkham, Unnat Pinsopon, Withit Chatlatanagulchai	
Paper ID: 1182 .....	153
<i>Attitude Error Cancellation for Strip-map SAR by Controlling Steering Angles of the Antenna Platform</i>	
Yasser Elkoteshy, Shuyuan Yang, and Fathy Abdelkader	
Paper ID: 1286 .....	154
<i>Hard Disk Drive Failure Mode Prediction based on Industrial Standard using Decision Tree Learning</i>	
Thanadon Suchatpong, Krischonme Bhumkittipich	
Paper ID: 1375 .....	155
<i>Optimizing PID Controller Tuning for Greenhouse Lighting Control System by Varying Number of Objectives</i>	
Mehdi Mahdavian, Naruemon Wattanapongsakorn	
<b>Digital Image Processing 1 .....</b>	<b>156</b>
Paper ID: 1056 .....	157
<i>"Combination" method of fast HEVC Encoding</i>	
Sapna Vasudevan, K. R. Rao	
Paper ID: 1157 .....	158
<i>Red Traffic Light Detection using Fast Radial Symmetry Transform</i>	
Sorn Sooksatra, Toshiaki Kondo	
Paper ID: 1263 .....	159
<i>Dual-Channel SAR Ground Moving Target Detection and Parameters Estimation</i>	
H. Shabana, Liu Fang, Shuyu Xiang	
Paper ID: 1269 .....	160
<i>Rice Growing Stage Monitoring in Small-scale Region Using ExG Vegetation Index</i>	
Narut Soontranon, Panu Srestasathien, Preesan Rakwatin	
Paper ID: 1312 .....	161
<i>Rice Cultivation and Harvest Date Identification Based on a Hidden Markov Model</i>	
Saran Suwannachatkul, Teerasit Kasetkasem, Kitti Chumkesornkulkit, Preesan Rakwatin, Thitiporn Chanwimaluang, Itsuo Kumazawa	
<b>Renewable Energy in Power System .....</b>	<b>162</b>
Paper ID: 1214 .....	163
<i>Effects of Fault Condition on the Performance of Hybrid PV/Wind Power Systems</i>	
Siavash Taghipour Broujeni, Seyed Hamid Fathi, Alireza Niknam Kumle	

# Hard Disk Drive Failure Mode Prediction based on Industrial Standard using Decision Tree Learning

Thanadon Suchatpong, Krischonme Bhumkittipich  
Department of Electrical Engineering, Faculty of Engineering  
Rajamangala University of Technology Thanyaburi,  
Pathumthani, Thailand  
Thanadon\_s@hotmail.com

**Abstract**— Quality and reliability are extremely important for Hard Disk Drive (HDD) manufacturer, which increasing upon the expansion capacity of HDD. The production process contains many tests in order to evaluate quality and efficiency which resulting high amount of complex data. Some HDDs will be randomly selected from each lot to test with various environments simulation. Due to complexity of tests and higher capacity of HDD manufacturer according to increasing demand from the consumer, for process test, it might take a long time. However, some failing HDD might be found at the customers test process or end users. This paper introduced the failure prediction using decision tree learning and the procedure of data collection and preparation. The main purpose of the study is to eliminate the simulation of the various environments in the tests which reducing time and increasing speed and efficiency of the overall analysis process prior to deliver the products to the customers. Moreover, Corrective Action: CA, the action that will solve and prevent the cause of the failure, can be verified faster. This study can improve quality and reliability of the production.

**Keywords**— Hard disk drive, Quality Checking, Pattern Classification

## I. INTRODUCTION

The hard disk drive is the most important thing in digital storage devices industry. In digital age, smart phone and mobile devices is the essential gadget for almost everyone in everyday life in order to share or transfer memories, activities which causing enormously data usage through the networks. Nowadays, a new technology that enabled the users to access sharing data and online activities faster called Cloud technology; it provides multiple sources of data storage server that means high capacity data storage device such hard disk drive, is necessary for the service providers. HDD has been developed consistently in the term of quality, capacity and data transfer rate. For example, to increase the areal density, the capacity per unit area of disk surface in HDD; many technologies will be applied which possibly causing higher sensitivity of internal mechanism and generating many new problems. Consequently, to guarantee the quality and increase the reliability of the products, Integration Failure Analysis is the important [1] procedure before delivered to the customers. Corrective action for problem prevention will be verified and determined from the relationship between the failure modes [2] and the root causes that analyzed upon drive's history and

bench test. Nevertheless, the extent of analysis is from the sampling HDDs.

Thus, this research uses data mining to classify failure mode for faster CA verification and deduction of analysis time as well as the failures in manufacturing process. Fig. 1, shows the manufacturing test process that is considered in this study. The process test is the main block consists of Pretest, Functional Test and Final Test. The Failure Analysis block is the process that analysis in term of electrical, mechanical and head analysis before failure mode conclusion to define further CA.

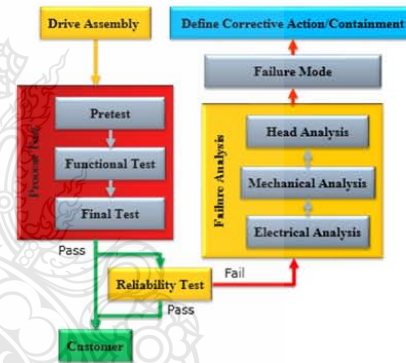


Fig. 1. Manufacturing test process.

The rest details of this paper after the first section are divided into the following sections: section 2 presents the methodology that consists of the failure mode, pre-processing and classification. The model and result are shown in section 3. Finally, section 4 gives the discussion and conclusion.

## II. METHODOLOGY

### A. Failure mode

The history of failure mode analysis from 1,161 HDDs of a new product can be summarized statistically the top 10 failure modes which are the following: 1) ATI/STE: a failing track is

interfered by writing operation at the adjacent track. 2) CND: the reporting problem/error is found in drive's history but not found or repeated during failure analysis 3) Degraded Head: reader/writer head performance degradation 4) FHM: resonance during read/write operation 5) Head Instability: reader/writer head performance unstable or intermittent that can cause by assembly process or contamination. 6) NMD: defect on disk that cannot be recovered by writer. 7) NPF: No any error is found 8) OTW: off-track write, the signal is written out of the path along the track. 9) VDW: vertical displaced write or skip write during write operation. 10) Weak Write: the written signal is weak.

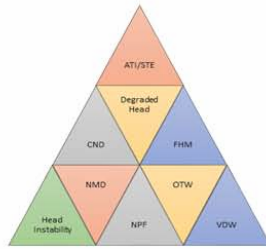


Fig. 2. Failure modes process.

For data preparation, we collected 1,161 failing HDDs from manufacturing process that was concluded by failure analysis and summarized into top 10 failure modes. Every HDD will be retrieved its value of each parameter as shown in Table I. The overall data showed different number of each failure mode as shown in Fig. 3 Generally, CA could fix only some failure modes permanently; the rest can be found repeatedly such as ATI/STE.

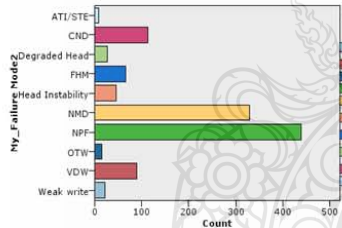


Fig. 3. Quantity of failure modes.

The significant parameters for each failure mode will be selected for considering in prediction (see Table I.) For example in Fig. 4 that shows the relationship between parameter 'RAW\_ERROR\_RATE' and the failure modes which helping classify the failure modes.

### B. Pre-processing

In order to improve the prediction accuracy; data pre-processing will be normalized by comparing the parameter value with threshold value that obtained from the test process experts and the threshold of each product is different from

each other. The conditions of normalization are: '1' or 'High' will be replaced if the parameter value is over its product threshold and '0' or 'Normal' will be replaced if the parameter value is under its product threshold. Data after normalization will be mixed [3] with discrete data and continuous data that can help in the classification.

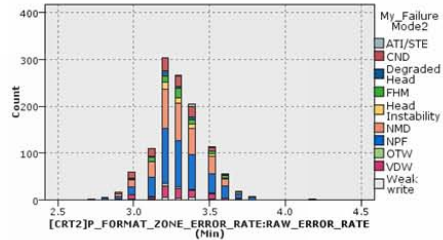


Fig. 4. Histogram of parameter RAW\_ERROR\_RATE with Failure Modes.

### C. Classification

After Pre-Processing, normalized data or training data will be tried to feed into many interesting algorithms such as 'Neural Net' [4] that classifies by generating a prediction through examining individual records, removing the weakest units in the hidden layers as training proceeds, 'C5.0' [5][6] that classifies by building a predictive model based on decision tree learning with boosting and pruning, 'Support Vector Machine' (SVM) [6] classifies by constructing a hyper plane that has the largest distance to the nearest training data point of any class in a high dimensional space, 'Classification and Regression' (C&R) Tree[7] classifies by building a predictive model based on decision tree learning with Regression, and 'CHAID' [8] classifies by building a predictive model based on decision tree learning with chi-squared distribution. The algorithms and test structure are shown in Fig. 5 with configuration in some feature to support with training data sets.

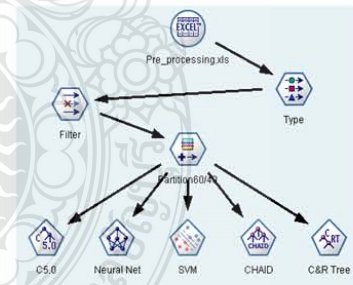


Fig. 5. Test structure and Algorithm.

C5.0 classification algorithm is selected in this study because it is supervised learning method that analyze data with multi class classifier. The algorithm is configured to enable the Boosting feature can be used [9] to improve decision tree

accuracy and Pruning feature for pruning the tree to minimize the decision trees.

Then, Cross-Validation will be used to split random data into 2 groups which are 60% of training sets for learning process in C5.0 algorithm and 40% test sets for testing data with decision tree model from training process. The data partition is shown in Fig. 6; the result shows accuracy improvement that is acceptable when using this model to predict the failure mode from manufacturing process.

TABLE I. PARAMETERS

Parameter	Description
Start Motor	Check Spin Time Mode.
Change motor speed	Spins drive at a higher than normal RPM for a period of time.
Servo Actuator Retract	Detect Servo Actuator retract failure.
Servo Seek	Determine maximum, minimum and average seek times of a drive using various cylinder and head modes.
Side Erasure Test	Measure the side track erasure effect of writing on a given track.
DC Thermal Asperity Scan	Scan Perform detection / burnishing / logging of asperities, and to merge Asperity Log entries into the Table.
Servo Flaw	Defective on servo wedges.
Contact Detect	Detect Head/Disk Contact for determine distance between Head and Disk.
Head Stability	Find unstable heads by writing/reading many times and reporting error counts for Timing Mark Not Detect, Track ID, Observer Sector Errors and Unsafe for each head.
Resonance	Measure the resonance frequencies and amplitudes on heads at the specified cylinder.
Skip Write Detect	Use the skip write detector to measure and adjust fly height.
Read Error Rate	Use a read scan of the target zone to determine the Read Error Rate.
Hot Cold Write Test	Check first sector and late sector writes, make sure read error rates are within limits at hot and cold temperature.
Primary Defect	Check the defect list length for validity and display the list of defects.
Zone Based Transfer Rate Test	Provide a mechanism to measure the delivered data rate over a large data transfer.
Bias Cal	Calibrate bias current and measure head resistance.
Write Current	Measure bias current on head in write operation.
AGC Data	Read the servo AGC setting.
OTC	Determine the off-track read capability after selectable write operations for the target track.
PES Data	Read Servo Position Error Signal and display the information read.
Delta VGA	Detect Dragging heads by using servo AGC measurements.
Scratch Fill	Identify possible scratches on the media.
SMART_Attribute	Verify the cumulative SMART attribute values on a drive are within spec from design center.

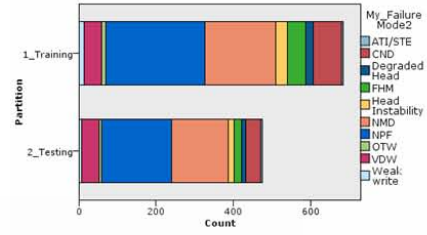


Fig. 6. Partition data.

Fig. 7 shows the manufacturing test process with failure mode prediction that is the propose algorithm. The failure mode prediction can be used effectively in the production process and also be developed feature to detect the signature of the problem and alert customer to backup data for unexpected damage of data that possibly occur in the future.

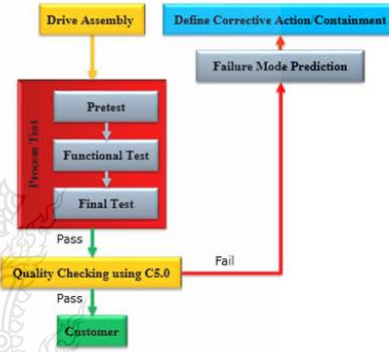


Fig. 7. Manufacturing test process with Failure Mode Prediction.

### III. RESULTS AND MODEL ESTABLISHED

In this paper, the total of 1,161 sampling HDDs were randomly selected from manufacturing process with a homogenous set. The data retrieved from the sampling HDDs will be normalize in data pre-processing then put through the C5.0 classification algorithm. The outputs were random split into 60% for training set and 40% for testing set. These findings were prepared with the results for each classification model, as shown in Fig. 5. It results that C5.0 model is the most accurate algorithm for prediction

The model of classification from C5.0 algorithm shows high accuracy for each failure modes and when compared with other algorithm at 98.54% for training and 99.58% for testing in Table III, because the data is compatible with the algorithm C5.0 which supporting multiple classes with discrete and continuous and has boosting and pruning algorithm inside in order to reduce tree size that caused higher accuracy. This model was created with cross-validation 10-fold that will be acceptable for using in the future.

According to the classification model verification, the important parameters in top 3 levels of decision tree as shown in Table IV. Its procedure is similar to failure analysis by the experts in order to classify the failures. Additionally, the new parameters that can help to classify the failure were found in this study. For example, "Degraded Head" and "Head Instability" can be classified using these following parameters: Contact Detect, Skip Write Detection, Primary Defect, Delta VGA, Write Current and Read Error Rate. In facts, "Degraded Head" and "Head Instability" cannot be classified using only Read Error Rate parameter but adding with the new multiple parameters such as Write Current for to classify which one is "Degraded Head" or "Head Instability" that shows the model is reliable to predict the failure mode in future.

Finally, automatic classification software will be developed from classification model to predict the failure mode during integration test in manufacturing process that can help to summarize the main failure symptoms of all drives for root cause investigation, corrective action verification and also protect some process escape drives delivered to the customers.

TABLE II. FAILURE MODE PREDICTION USING C5.0 ALGORITHM

Failure Mode	Correct	Wrong	Accuracy
ATI/STE	4	0	100%
CND	40	0	100%
Degraded Head	9	1	90%
FHM	19	0	100%
Head Instability	16	0	100%
NMD	146	0	100%
NPF	182	0	100%
OTW	6	0	100%
VDW	44	1	97.78%
Weak Write	8	0	100%

TABLE III. MODEL RESULTS WITH THE SAME DATA SET THROUGH DIFFERENT

Algorithm	Training			Testing		
	Correct	Wrong	Accuracy	Correct	Wrong	Accuracy
C5.0	675	10	98.54%	474	2	99.58%
Nueral Net	337	348	49.2%	267	209	56.09%
SVM	269	416	39.27%	184	292	38.66%
CHAID	363	322	52.99%	240	236	50.42%
C&R Tree	344	341	50.22%	271	205	56.93%

#### IV. CONCLUSIONS

This paper has shown C5.0 algorithm can be used for multi-class classifier with mixing types of data and shows high accuracy at 99.58% for predicting failure mode from test parameter that is an acceptable model for using to develop

software to classify data from test parameter automatically when pass test process from manufacturing process. It will help to reduce the time of analysis and expert engineer can grouping and sampling the represent HDDs from each failure mode to verify and do failure analysis for finding CA faster and take some actions to manufacturing process to improve quality and reliability and can stop sending defective HDDs to the customers. Automatic software can be developed to analyze data for manufacturing process before delivered the HDDs to the customers, and also for increasing quality and reliability for production and customer.

TABLE IV. TOP 3 LEVEL OF DECISION TREE FROM TOP 5 RULES

Rules	Root node	Node Lv.2	Level 3
1	Contact Detect	Servo Flaw	Scratch Fill Read Error Rate PES Data
		Delta VGA	Resonance
2	Servo Flaw	Read Error Rate	Contact Detect Delta VGA
		Delta VGA	Resonance OTC
3	Error Rate	Skip Write Detect	AGC Data Write Current PES Data
		Resonance	Contact Detect
4	Contact Detect	Skip Write Detect	Write Current Bias Cal
		Delta VGA	Scratch Fill Primary Defect OTC
5	Delta VGA	Resonance	PES Data Scratch Fill
		AGC Data	Primary Defect

#### REFERENCES

- [1] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado, "Improved disk drive failure warnings," in IEEE transactions on reliability, September 2002.
- [2] R. Tanapong, P. Jiradej, "Hard disk drive failure mode prediction for Head problem from SMART attributes using Data mining method, JCSE 2011, Nakhon Pathom, Thailand, 2011.
- [3] R. Tanapong, P. Jiradej, "Hard Disk Drive Failure Mode Prediction from SMART Attribute using Data Mining Method", DST-CON 2011, Bangkok, Thailand, 2011.
- [4] T. Waraporn, K. Yuttana, "Customer Failure Modes Prediction for Hard Disk Drives using Neural Networks Rank-Level Fusion", ECTICON 2011, Khon Kaen, Thailand, p.476.
- [5] W. Meng, G. Kun, "A Novel Hyperspectral Classification Method Based on C5.0 Decision Tree of Multiple Combined Classifiers", ICCIS 2012, Chongqing, China, 2012.
- [6] P. Pongpatarakon "A Comparative Study of Classification Properties between CART, SVM, C5.0 and Hybrid Methods", NCCIT 2009, Bangkok, Thailand, 2009.
- [7] M. Garg, U. Singh, "C&R Tree based Air Target Classification Using Kinematics", IJCCT 2012, Vol3 Issue 1, 2012.
- [8] M. Turea, I. Kurta, A. Turhan Kurumb, "Comparing classification techniques for predicting essential hypertension", Expert Systems with Applications 29 (2005), p.583-588.
- [9] A. M. Elsayad "Diagnosis of Breast Tumor using Boosted Decision Trees" ICGST-AIML Journal, Volume 10, Issue 1, October 2010.

## ประวัติผู้เขียน

ชื่อ-สกุล	นายธนดล สุชาติพงศ์
วัน เดือน ปีเกิด	24 มิถุนายน 2529
ที่อยู่	234/6 หมู่ที่ 8 ต.ชะมาย อ.ทุ่งสง จ.นครศรีธรรมราช 80110
การศึกษา	ปริญญาตรี คณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ปี พ.ศ. 2551
ประสบการณ์การทำงาน	วิศวกร ฝ่ายวิเคราะห์และตรวจสอบคุณภาพผลิตภัณฑ์ บริษัท ซีเกท เทคโนโลยี (ประเทศไทย) จำกัด พ.ศ. 2552 ถึงปัจจุบัน
เบอร์โทรศัพท์	08-9645-4388
อีเมลล์	Thanadon_s@hotmail.com

