แบบจำลองเพื่อพัฒนาคุณภาพของผลิตภัณฑ์เอชจีเอในโรงงานอุตสาหกรรมฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ

นางสาวอโณทัย ศิลเทพาเวทย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository(CUIR)
are the thesis authors' files submitted through the Graduate School.

A MODEL FOR IMPROVING QUALITY OF PRODUCT HGA IN HARDDISK MANUFACTURING BY DECISION TREE TECHNIQUE

Miss Anotai Siltepavet

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	แบบจำลองเพื่อพัฒนาคุณภาพของผลิตภัณฑ์เอชจีเอในโรงงานอุตสาหกรรมฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ
โดย	นางสาวอโณทัย ศิลเทพาเวทย์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

…… ประธานกรรมการ
(ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

…… กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.วรเศรษฐ สุวรรณิก)

อโณทัย ศิลเทพาเวทย์ : แบบจำลองเพื่อพัฒนาคุณภาพของผลิตภัณฑ์เอชจีเอในโรงงานอุตสาหกรรมฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ. (A MODEL FOR IMPROVING QUALITY OF PRODUCT HGA IN HARDDISK MANUFACTURING BY DECISION TREE TECHNIQUE) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.สุกรี สินธุภิญโญ, 62 หน้า.

วิทยานิพนธ์เล่มนี้เสนอการสร้างแบบจำลองเพื่อพัฒนาคุณภาพการผลิตฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ แบบจำลองนี้ช่วยบอกแนวทางการปรับพารามิเตอร์ในโรงงานให้ได้จำนวนงานที่ดีมากที่สุด วิธีการแบ่งออกเป็นสองส่วน ส่วนแรกระบุพารามิเตอร์ ที่สำคัญในโรงงานซึ่งมีผลต่อคุณภาพของชิ้นงานด้วยต้นไม้ตัดสินใจ ส่วนที่สองเป็นการปรับพารามิเตอร์โดยการปรับแต่ละพารามิเตอร์นั้นจะต้องไม่ส่งผลกระทบต่อพารามิเตอร์อื่นๆ โดยใช้ค่าสหสัมพันธ์และสมการเชิงเส้นถดถอยมาใช้ในการหาการเปลี่ยนแปลงเมื่อเกิดการปรับพารามิเตอร์ ในงานวิจัยนี้ทำการทดลองปรับพารามิเตอร์ 3 วิธี คือ ปรับจากพารามิเตอร์จากบน-ล่าง ล่าง -บน และแบบผสมผสาน ทำการเลือกวิธีการปรับที่ทำให้สามารถเพิ่มจำนวนชิ้นงานได้จำนวนมากที่สุดมาใช้ในการสร้างแบบจำลองในการพัฒนากระบวนการผลิตฮาร์ดดิสก์ต่อไป

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต.....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2554.....

5371469021: MAJOR COMPUTER SCIENCE

KEYWORDS : DECISION TREE / HARD DRIVE / IMPROVE QUALITY OF PRODUCT

ANOTAI SILTEPAVET : A MODEL FOR IMPROVING QUALITY OF PRODUCT HGA IN HARDDISK MANUFACTURING BY DECISION TREE TECHNIQUE. ADVISOR : ASST.PROF. SUKREE SINTHUPINYO, 62 pp.

This paper is aimed at constructing a model which can improve quality of products in hard drive manufacturing by Decision Tree Learning technique. The output model will be a parameter adjustment guideline in the development-process. It guides the best method to adjust controllable parameters that take the lowest cost. The analysis of this proposed model is divided into two parts. The first part provides an analysis on a decision tree that can determine the parameters which affect the quality of the work-pieces. In the second part, the adjustability of the parameter is discussed and analyzed whether it can be adjusted without affecting other parameters based on the correlation coefficient and linear regression techniques. Furthermore, three case studies are illustrated to identify the adjustment methods, i.e. the Top-down, Bottom-up, and Hybrid method, then the best method which covers all possible value of adjustable parameters will be selected.

Department : Computer Engineering Student's Signature .....

Field of Study : Computer Science Advisor's Signature .....

Academic Year : 2011 .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จเรียบร้อยได้ด้วยดีเพราะได้รับคำแนะนำและให้คำปรึกษาจากศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา และ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ ซึ่งเป็นผู้ชี้แนะแนวทางในการศึกษาและทำการวิจัยจนเกิดเป็นวิทยานิพนธ์ฉบับนี้ขึ้น นอกจากนี้ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ ซึ่งได้แก่ ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา ผู้ซึ่งเป็นประธาน ผู้ช่วยศาสตราจารย์ ดร.วรเศรษฐ สุวรรณิก ผู้ซึ่งเป็นกรรมการที่ได้สละเวลาอันมีค่ามาใช้ในการชี้แจงถึงข้อบกพร่อง รวมถึงแนวทางการแก้ไขและข้อแนะนำดีๆ ให้แก่ข้าพเจ้า

ขอขอบพระคุณ คุณพ่อ คุณแม่ รวมไปถึง พี่ๆ ที่ทำงาน และ เพื่อนๆ ที่คอยให้คำแนะนำและกำลังใจดีๆ เมื่อข้าพเจ้าพบปัญหาต่างๆ ในขณะทำงานวิจัยเล่มนี้ ทำให้ข้าพเจ้ามีพลังที่จะต่อสู้และฝ่าฟันอุปสรรคต่างๆ เพื่อให้วิทยานิพนธ์เล่มนี้สำเร็จสมบูรณ์ไปด้วยดี และยิ่งไปกว่านั้นต้องขอขอบคุณต่อเจ้าหน้าที่ของภาควิศวกรรมศาสตร์คอมพิวเตอร์ทุกท่านที่อำนวยความสะดวกในทุกๆ เรื่องที่เกี่ยวกับการศึกษา

สุดท้ายนี้ ขอขอบพระคุณทุกท่านอีกครั้งที่คอยส่งเสริมและให้กำลังใจ รวมถึงคำปรึกษาที่ดี ในการศึกษาและทำงานวิจัยให้ประสบความสำเร็จในจุฬาลงกรณ์มหาวิทยาลัยแห่งนี้

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฎ

บทที่ 1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

วัตถุประสงค์ของงานวิจัย..... 3

ขอบเขตของการวิจัย..... 3

ประโยชน์ที่คาดว่าจะได้รับ..... 4

ขั้นตอนการดำเนินการวิจัย..... 4

โครงสร้างของวิทยานิพนธ์..... 4

ผลงานที่ตีพิมพ์จากวิทยานิพนธ์..... 5

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 6

2.1 แนวคิดและทฤษฎี..... 6

2.1.1 กระบวนการผลิตฮาร์ดดิสก์..... 6

2.1.2 ต้นไม้ตัดสินใจ (Decision Tree)..... 9

2.1.2 Weka..... 14

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 14

บทที่ 3 ระเบียบขั้นตอนวิธีที่นำเสนอ..... 25

ระเบียบขั้นตอนวิธี..... 25

3.1 การทำความสะอาดข้อมูล..... 27

3.2 การแบ่งข้อมูล..... 27

หน้า

3.3 การสร้างต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5..... 28
3.4 การปรับพารามิเตอร์แบบ บนลงล่าง ล่างขึ้นบน และแบบผสม..... 28
3.4.1 การปรับพารามิเตอร์แบบบนลงล่าง..... 29
3.4.2 การปรับพารามิเตอร์แบบล่างขึ้นบน..... 31
3.4.3 การปรับพารามิเตอร์แบบผสม..... 34
บทที่ 4 การทดลองและผลการทดลอง..... 36
4.1 เครื่องมือที่ใช้..... 36
4.2 ข้อมูลที่นำมาใช้ในการทดลอง..... 36
4.3 การทดลองปรับพารามิเตอร์ของ product1..... 37
4.4 การทดลองปรับพารามิเตอร์ของ product2..... 43
4.5 การทดลองปรับพารามิเตอร์ของ product3..... 52
4.6 วิเคราะห์ผลการทดลอง..... 58
4.6 สรุปผลการทดลอง..... 59
บทที่ 5 ข้อเสนอแนะและแนวทางการพัฒนาต่อ..... 60
รายการอ้างอิง..... 61
ประวัติผู้เขียนวิทยานิพนธ์..... 62

สารบัญตาราง

หน้า

ตารางที่ 2.1 ตารางการทดลอง FDS EXPERIMENTAL RESULTS..... 21
ตารางที่ 4.1 ข้อมูลการผลิตฮาร์ดดิสก์ที่นำมาใช้ในการทดลอง..... 36
ตารางที่ 4.2 การปรับข้อมูลของ Product 1 แบบบนลงล่าง..... 41
ตารางที่ 4.3 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 1..... 41
ตารางที่ 4.4 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 2..... 42
ตารางที่ 4.5 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 3..... 42
ตารางที่ 4.6 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 4..... 42
ตารางที่ 4.7 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 1 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20..... 43
ตารางที่ 4.8 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 1 ด้วยชุดข้อมูลทดสอบ..... 43
ตารางที่ 4.9 การปรับข้อมูลของ Product 2 แบบบนลงล่าง..... 48
ตารางที่ 4.10 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 1..... 49
ตารางที่ 4.11 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 2..... 49
ตารางที่ 4.12 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 3..... 50
ตารางที่ 4.13 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 4..... 50
ตารางที่ 4.14 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 5..... 50
ตารางที่ 4.15 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 6..... 51
ตารางที่ 4.16 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 2 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20..... 51
ตารางที่ 4.17 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 2 ด้วยชุดข้อมูลทดสอบ..... 52
ตารางที่ 4.18 การปรับข้อมูลของ Product 3 แบบบนลงล่าง..... 56

หน้า

ตารางที่ 4.19 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 1..... 56
ตารางที่ 4.20 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 2..... 56
ตารางที่ 4.21 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 3..... 57
ตารางที่ 4.22 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 4..... 57
ตารางที่ 4.23 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 3 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20..... 58
ตารางที่ 4.24 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 3 ด้วยชุดข้อมูลทดสอบ..... 58

สารบัญภาพ

		หน้า
ภาพที่ 2.1	ลำดับขั้นตอนการผลิตฮาร์ดดิสก์	8
ภาพที่ 2.2	ผังงานของการวิเคราะห์ปัญหา	14
ภาพที่ 2.3	ผังงานของโมเดลการวิเคราะห์วิเคราะห์ความผิดพลาด	16
ภาพที่ 2.4	ตัวอย่างต้นไม้ตัดสินใจที่ได้	18
ภาพที่ 2.5	แผนผังการทำงานโดยรวมของ FD โมเดล	19
ภาพที่ 2.6	แผนผังการสร้าง FD model	20
ภาพที่ 2.7	Flow การทำงาน ด้วยวิธี Sampling-based sensitivity	22
ภาพที่ 3.1	ขั้นตอนการพัฒนาคุณภาพการผลิตฮาร์ดดิสก์	25
ภาพที่ 3.2	วิธีการแบ่งข้อมูลในการทดลอง	26
ภาพที่ 3.3	สัดส่วนการแบ่งข้อมูลการผลิต	27
ภาพที่ 3.4	อัลกอริทึมบนลงล่าง	29
ภาพที่ 3.5	ตัวอย่างต้นไม้ตัดสินใจ1	30
ภาพที่ 3.6	อัลกอริทึมล่างขึ้นบน	32
ภาพที่ 3.7	ตัวอย่างต้นไม้ตัดสินใจ2	33
ภาพที่ 3.8	อัลกอริทึมแบบผสม	34
ภาพที่ 4.1	สัดส่วนของในการแบ่งข้อมูลของ Product 1	37
ภาพที่ 4.2	รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 1	38
ภาพที่ 4.3	ภาพต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 1	38
ภาพที่ 4.4	รายงานต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 1	39
ภาพที่ 4.5	ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 1	39
ภาพที่ 4.6	รายงานต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 1	40
ภาพที่ 4.7	ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 1	40
ภาพที่ 4.8	สัดส่วน ในการแบ่งข้อมูลของ Product 2	44
ภาพที่ 4.9	รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 2	45

หน้า

ภาพที่ 4.10 ต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 2..... 45
ภาพที่ 4.11 รายงานต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 2..... 46
ภาพที่ 4.12 ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 2..... 46
ภาพที่ 4.13 รายงานต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 2..... 47
ภาพที่ 4.14 ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 2..... 47
ภาพที่ 4.15 สัดส่วนของ Product 3 ในการแบ่งข้อมูลของ Product 3..... 52
ภาพที่ 4.16 รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 3..... 53
ภาพที่ 4.17 ต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ 80 ของ Product 3..... 53
ภาพที่ 4.18 รายละเอียดต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 3..... 54
ภาพที่ 4.19 ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 ของ Product 3..... 54
ภาพที่ 4.20 รายละเอียดต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 3..... 55
ภาพที่ 4.21 ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ Product 3..... 55

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

อุตสาหกรรมการผลิตฮาร์ดดิสก์เป็นอุตสาหกรรมที่สร้างรายได้เข้าประเทศเป็นอย่างมาก [1] เนื่องจากประเทศไทยเป็นผู้ส่งออกฮาร์ดดิสก์รายใหญ่ของโลก ดังนั้น การพัฒนาคุณภาพอุตสาหกรรมการผลิตฮาร์ดดิสก์ไทยจึงเป็นสิ่งสำคัญที่จะทำให้ประเทศไทยยังคงเป็นประเทศที่มีส่วนแบ่งทางตลาดมากที่สุด

กระบวนการผลิตฮาร์ดดิสก์แบ่งเป็น 5 กระบวนการหลักต่อเนื่องตามลำดับคือ

1. แผ่นเวเฟอร์ หรือ Wafer Fabrication คือการเทสารเป็นชั้นๆ ด้วยสารเคมีที่มีคุณสมบัติเป็นสารแม่เหล็กซึ่งช่วยในการเขียนและอ่านข้อมูลจากจานหมุนได้
2. หัวอ่าน และ เขียนข้อมูล หรือ Slider Fabrication คือการนำแผ่นเวเฟอร์จากกระบวนการผลิตแผ่นเวเฟอร์ มาทำการตัดเป็นชิ้นๆ ซึ่งแต่ละชิ้นคือหัวอ่านและเขียนข้อมูล 1 หัว แต่ละชิ้นต้องมีการทำโครงหน้าขึ้นมาเพื่อช่วยในการทรงตัวของหัวอ่านเขียนในขณะฮาร์ดดิสก์ทำงานจริง
3. Head Gimbal Assembly หรือ HGA คือการนำหัวหัวอ่านและเขียนข้อมูลที่ได้จากกระบวนการผลิตหัวเขียนอ่านข้อมูล มาประกอบเข้ากับแกนของหัวอ่านและเขียนข้อมูล ซึ่งในกระบวนการนี้จะมีการทดสอบทางไฟฟ้าว่าเมื่อชิ้นส่วนประกอบกันแล้วประสิทธิภาพของหัวหัวอ่านและเขียนข้อมูลเป็นอย่างไร
4. Head Stack Assembly หรือ HSA คือ การนำหัวอ่านที่ประกอบแกนแล้วจากกระบวนการผลิต Head Gimbal Assembly มาประกอบรวมกัน โดยใน 1 HSA อาจมีได้หลาย HGA หรือ หลายแกนและหลายหัวอ่านได้ ขึ้นอยู่กับขนาดความจุของฮาร์ดดิสก์นั้น
5. Hard Disk Drive หรือ HDD คือการนำ HSA มาประกอบกับตัว Drive และทดสอบทั้งชุดเป็นการเสร็จสิ้นกระบวนการผลิต

งานวิจัยนี้นำเสนอแนวทางที่จะช่วยในการพัฒนาคุณภาพอุตสาหกรรมการผลิตฮาร์ดดิสก์ด้วยการออกแบบโมเดลที่ใช้ในการกำหนดเงื่อนไขของปัจจัยที่มีผลต่อการผลิตของผู้ขายวัตถุดิบแก่โรงงานซึ่งในทุกขั้นตอนของกระบวนการผลิตจะต้องมีการทดสอบวัดค่าพารามิเตอร์เพื่อนำค่าพารามิเตอร์ที่ได้เหล่านี้มาวิเคราะห์จำแนกประเภทชิ้นงานนั้นๆว่าเป็นงานดีหรืองานเสีย โดยจำนวนชิ้นงานที่ดีต่อจำนวนงานทั้งหมดที่ผลิต เรียกว่า “Yield” การทำปริมาณผลผลิตในการผลิตให้สูงที่สุดคือเป้าหมายหลักของโรงงาน เพราะ ปริมาณผลผลิตจะเป็นค่าที่บ่งบอกคุณภาพของสินค้า และ ขั้นตอนการควบคุมกระบวนการผลิตของโรงงาน ซึ่งในที่นี้ คือ การกำหนดข้อกำหนดของสินค้า ดังนั้น การควบคุมการผลิตโดยการปรับค่าพารามิเตอร์ให้อยู่ในข้อกำหนดจะทำให้งานที่ผลิตได้งานดีออกมาจำนวนมากที่สุด

ปัจจุบันนักวิเคราะห์การผลิตของโรงงานต้องทำการตรวจสอบตัวอย่างที่มีผลต่อชิ้นงานอย่างสม่ำเสมอและปรับข้อกำหนดของแต่ละพารามิเตอร์เพื่อให้ได้ค่าที่เหมาะสมเพื่อปรับปรุงกระบวนการผลิตให้เพิ่มจำนวนชิ้นงานคุณภาพดี และบรรลุเป้าหมายหลักของโรงงานคือปริมาณผลผลิตที่สูง ค่าพารามิเตอร์ต่างๆ ที่ได้จากการทดสอบในการผลิตถูกเก็บไว้ในคลังข้อมูล (Data Warehouse) และเมื่อนำมาทดสอบข้อกำหนดจะสามารถจำแนกประเภทของชิ้นงานที่ผลิตได้ ว่าเป็นงานดีหรืองานเสีย แต่ทั้งนี้ทั้งนั้นข้อกำหนดนั้นไม่ได้คงที่ตลอดเวลาสามารถเปลี่ยนแปลงได้ เมื่อสภาวะการผลิตมีการเปลี่ยนแปลงไป เช่น ผลิตภัณฑ์นั้นๆ ต้องมีการเปลี่ยนวัตถุดิบ หรือเปลี่ยนกระบวนการต่างๆ เมื่อผลิตไปได้ระยะหนึ่ง หรือ เกิดผลิตภัณฑ์ใหม่ที่ยังไม่สามารถกำหนดข้อกำหนดที่แน่ชัดได้เนื่องจากข้อมูลการผลิตมีน้อย นักวิเคราะห์การผลิตของโรงงานจึงต้องปรับเปลี่ยนข้อกำหนดตลอดเวลาเพื่อให้โรงงานผลิตสินค้าที่มีคุณภาพมากขึ้น แต่การทำงานจริงนั้นต้องใช้เวลานาน เนื่องจากนักวิเคราะห์การผลิตของโรงงานใช้วิธีการดึงข้อมูลมาเปรียบเทียบหาความสัมพันธ์ของแต่ละพารามิเตอร์ ด้วยกราฟทางสถิติเพื่อหากลุ่มพารามิเตอร์ที่สำคัญ ซึ่งมีขั้นตอนที่ยุ่งยาก

งานวิจัยนี้ได้เสนอการออกแบบจำลองที่ช่วยเพิ่มชิ้นงานดีด้วยการปรับพารามิเตอร์ที่สามารถปรับได้ในโรงงาน เพื่อช่วยในการตัดสินใจของนักวิเคราะห์ผลผลิตของโรงงานและลดเวลาการทำงาน ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision Tree) และ ปรับแต่ง

ต้นไม้ด้วยการหาค่าสหสัมพันธ์ ระหว่างพารามิเตอร์เพื่อใช้ในการออกแบบการตั้งค่าข้อกำหนดของพารามิเตอร์ที่เหมาะสม โดยพารามิเตอร์นั้นสามารถปรับค่าได้เพื่อช่วยเพิ่มงานดีในการผลิตงานในโรงงานให้มากขึ้น โดยทำการศึกษาข้อมูลจากโรงงานผลิตฮาร์ดดิสก์ในส่วนกระบวนการผลิต Head Gimbal Assembly หรือ HGA ซึ่งเป็นกระบวนการที่มีผลสืบเนื่องจาก ขั้นตอนกระบวนการผลิตผ่านเวเฟอร์ หรือ Wafer Fabrication และ กระบวนการผลิตหัวอ่าน และ เขียนข้อมูล หรือ Slider Fabrication ทั้งสามส่วนนี้เป็นกระบวนการที่เกี่ยวกับหัวเขียนอ่าน ซึ่งเป็นส่วนประกอบสำคัญที่สุดของฮาร์ดดิสก์ ข้อมูลที่นำมาใช้นั้นประกอบไปด้วย ค่าพารามิเตอร์ต่างๆที่ได้จากการทดสอบหลังขั้นตอนการผลิต และ ค่าที่บอกประเภทของงานที่ผลิตว่าเป็นงานดี หรืองานเสีย มาทำการเรียนรู้ด้วยเทคนิคต้นไม้ตัดสินใจเพื่อจำแนกความสำคัญของค่าพารามิเตอร์ต่างๆ ว่ามีผลกระทบต่อปริมาณผลผลิตอย่างไร ควรแก้ไขที่จุดไหนเมื่อพบปัญหา และ เกณฑ์ระดับอย่างต่ำที่ควรกำหนดสำหรับพารามิเตอร์นั้นๆ และ เทคนิคทางสถิติในการหาขอบเขตของพารามิเตอร์ที่สามารถขยายได้มากที่สุดโดยไม่กระทบต่อพารามิเตอร์อื่นๆ

วัตถุประสงค์ของการวิจัย

เพื่อหาวิธีการปรับพารามิเตอร์ที่ดีที่สุดสำหรับโรงงานผลิต Head Gimbal Assembly หรือ HGA เพื่อให้ได้จำนวนชิ้นงานที่มีคุณภาพดีมากขึ้น

ขอบเขตงานวิจัย

1. ใช้ข้อมูลการผลิตฮาร์ดดิสก์ 3 ชนิดในโรงงาน
2. ปรับค่าพารามิเตอร์เฉพาะพารามิเตอร์ที่สามารถปรับค่าได้ โดยใช้วิธีการถดถอยเชิงเส้น(Linear Regression) เพื่อปรับค่าพารามิเตอร์อื่นๆ ที่ขึ้นกับค่าพารามิเตอร์ที่ปรับค่าได้เหล่านั้น
3. ใช้ข้อมูลจากชิ้นงานรวมทั้งสิ้น 115,363 ชิ้นงาน ซึ่งเป็นข้อมูลระหว่าง เดือน มีนาคม 2554 ถึง เดือน มีนาคม 2555

4. เนื่องจากในการวิจัยครั้งนี้ ไม่สามารถทดลองปรับพารามิเตอร์ต่างๆ ได้จริง อันเนื่องมาจากประเด็นความเสี่ยงในการผลิตจริง ผู้วิจัยจึงวัดผลการทดลองจากกลุ่มข้อมูลจากข้อมูลทั้งหมด โดยแบ่งไว้ 50% สำหรับข้อมูลเรียนรู้และการปรับพารามิเตอร์ตามต้นไม้ตัดสินใจ และอีก 50% ที่เหลือ นำไปสร้างเป็นข้อมูลทดสอบการปรับพารามิเตอร์

ประโยชน์ที่คาดว่าจะได้รับ

วิทยานิพนธ์นี้นำเสนอวิธีการที่สามารถช่วยในการพัฒนาการผลิตทางด้านอุตสาหกรรม เป็นเครื่องมือช่วยบุคลากร ลดเวลาการทำงาน ลดต้นทุนการผลิต และลดของเสียในโรงงาน

ขั้นตอนการดำเนินงานวิจัย

1. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2. ศึกษาขั้นตอนการทำงานของนักวิเคราะห์การผลิตของโรงงาน ในการปรับปรุงคุณภาพผลิตภัณฑ์
3. ออกแบบวิธีการปรับพารามิเตอร์โดยใช้ต้นไม้ตัดสินใจ
4. ทำการทดลองตามวิธีที่นำเสนอ
5. วิเคราะห์ผลการทดลอง
6. สรุปผลการวิจัย และเรียบเรียงวิทยานิพนธ์

โครงสร้างของวิทยานิพนธ์

โครงสร้างวิทยานิพนธ์ฉบับนี้ประกอบไปด้วย 5 บทหลักคือ บทนำ เอกสารและทฤษฎีที่เกี่ยวข้อง ระเบียบขั้นตอนวิธีที่เสนอ การทดลองกับผลการทดลอง และบทสรุปกับแนวทางในการพัฒนาต่อ

ในบทแรกจะกล่าวถึง ประวัติความเป็นมา ปัญหา วัตถุประสงค์ ขอบเขตของงานวิจัย ประโยชน์ที่จะได้รับ ขั้นตอนการทำวิจัย โครงสร้างของวิทยานิพนธ์ และผลงานที่ตีพิมพ์

จากวิทยานิพนธ์ บทที่สอง อธิบายถึง ทฤษฎีต้นไม้การตัดสินใจ และ งานวิจัยที่เกี่ยวข้อง บทที่สาม กล่าวถึงระเบียบขั้นตอนวิธีที่เสนอ บทที่4 อธิบายผลการทดลองและวิเคราะห์ผลการทดลองที่ได้ และสุดท้ายในบทที่ห้า จะกล่าวบทสรุปที่ได้จากงานวิจัย และแนวทางในการพัฒนางานวิจัยนี้ต่อเพื่อปรับปรุงให้งานวิจัยนี้มีประสิทธิภาพครอบคลุมปัญหาต่างๆ ยิ่งขึ้นต่อไป

ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตอบรับให้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “A Heuristic Approach to Improve The Decision Tree of Product Acceptance in Hard Drive Manufacturing” โดย อโณทัย ศิลเทพาเวทย์ สุกรี สินธุภิญโญ และ ประภาส จงสถิตย์วัฒนา ในการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 8 ที่ประเทศไทยระหว่างวันที่ 9-10 พฤษภาคม 2555

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะอธิบายถึงแนวคิดและทฤษฎี รวมไปถึงเอกสารงานวิจัยที่เกี่ยวข้องกับงานวิจัยที่นำเสนอ เพื่อให้เห็นถึงพื้นหลังของงานวิจัยและปัญหา และแนวทางแก้ไขปัญหาในแบบต่างๆ

2.1 แนวคิดและทฤษฎี

ในหัวข้อนี้จะเป็นการอธิบายถึงกระบวนการผลิตฮาร์ดดิสก์ การเรียนรู้ของเครื่อง (Machine Learning) ด้วยต้นไม้ตัดสินใจ (Decision Tree) และเครื่องมือที่ใช้ในการทดลองซึ่งเป็นเครื่องมือเกี่ยวกับการเรียนรู้ของเครื่อง

2.1.1 กระบวนการผลิตฮาร์ดดิสก์

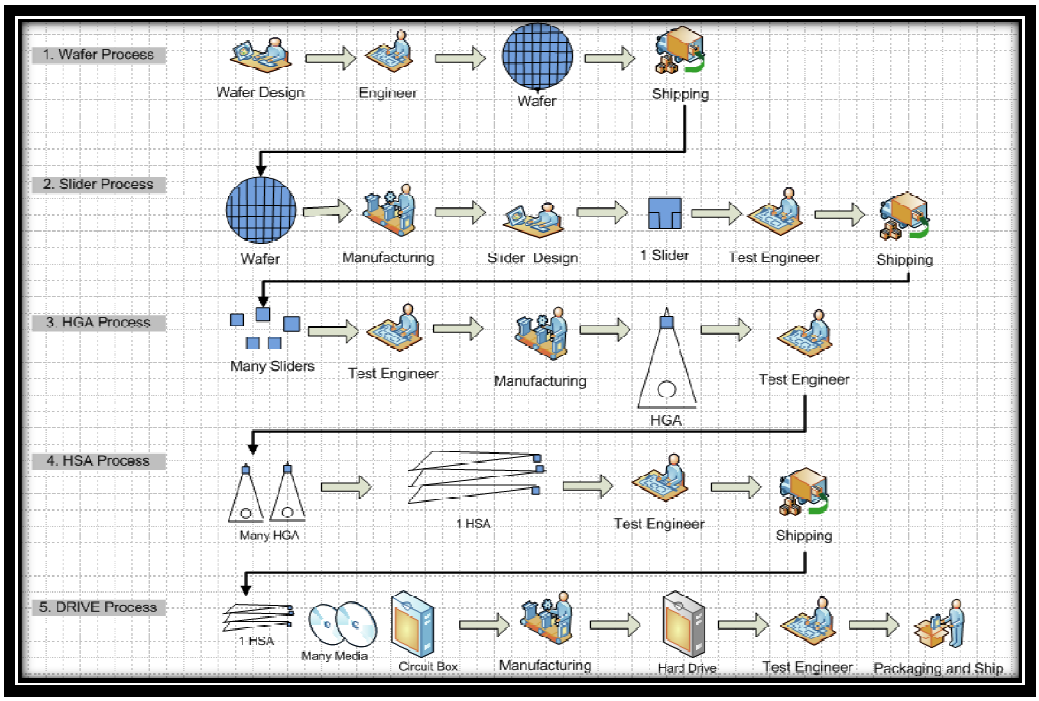
ฮาร์ดดิสก์เป็นอุปกรณ์ที่สำคัญในระบบคอมพิวเตอร์เพราะใช้เป็นแหล่งเก็บข้อมูลต่างๆ ทั้งโปรแกรมและข้อมูลของผู้ใช้ ในปัจจุบันมีรูปแบบของสื่อสารสนเทศต่างๆ มากมายซึ่งความละเอียดคุณภาพมากทำให้ขนาดของไฟล์จะมากตามด้วย ดังนั้นแหล่งจัดเก็บข้อมูลอย่างฮาร์ดดิสก์จึงจำเป็นต้องเพิ่มความจุให้มากขึ้นเรื่อยๆ ดังนั้นกระบวนการผลิตฮาร์ดดิสก์จึงถูกทำการพัฒนาอย่างต่อเนื่องเพื่อใช้รองรับโปรแกรมและข้อมูลจำนวนที่มากขึ้นโดยที่ยังสามารถอ่านเขียนข้อมูลได้อย่างรวดเร็ว ซึ่งการผลิตฮาร์ดดิสก์ที่มีคุณภาพหนึ่งตัวมีขั้นตอนการผลิตและการตรวจสอบต่างๆ ดังนี้

กระบวนการผลิตฮาร์ดดิสก์แบ่งเป็น 5 กระบวนการหลักต่อเนื่องตามลำดับคือ

1. กระบวนการผลิตแผ่นเวเฟอร์ หรือ Wafer Fabrication คือการเทสารเป็นชั้นๆ คล้ายชั้นของเวเฟอร์ด้วยสารเคมีที่มีคุณสมบัติเป็นสารกึ่งตัวนำซึ่งมีส่วนช่วยในการเขียนและอ่านข้อมูลจากจานหมุนได้ขั้นตอนการเทสารในแต่ละชั้นมีวิธีที่แตกต่างกันออกไปและสารที่ใช้ในแต่ละชั้นขึ้นอยู่กับการออกแบบรูปร่างของหัวอ่านเขียน ซึ่งแบบแต่ละอย่างจะให้ประสิทธิภาพการเขียนและอ่านข้อมูลที่แตกต่างกันออกไป ดังนั้นการออกแบบแผ่นเวเฟอร์ที่ดีจึงเป็นขั้นตอนที่สำคัญมากขั้นตอนหนึ่ง

2. กระบวนการผลิตหัวอ่าน และ เขียนข้อมูล หรือ Slider Fabrication คือการนำแผ่นเวเฟอร์จากกระบวนการ ผลิตแผ่นเวเฟอร์ มาทำการตัดเป็นชิ้นๆ ซึ่งแต่ละชิ้นคือหัวอ่านเขียน 1 หัว แผ่นเวเฟอร์เมื่อถูกนำมาที่กระบวนการนี้ จะถูกทำการแปรรูป ด้วยการขัด การกัดเซาะ เพื่อให้ได้รูปร่างหน้าตามการออกแบบของหัวอ่านเขียนในแต่ละรุ่นของฮาร์ดดิสก์ซึ่งรูปร่างหน้าของหัวอ่านเขียนนั้น มีผลต่อประสิทธิภาพการอ่านเขียนจริง ซึ่งเกิดจากการยกตัวของหัวอ่านเขียนในขณะที่อยู่บนแผ่นจานหมุนเมื่อใช้งานจริง
3. กระบวนการผลิต Head Gimbal Assembly หรือ HGA คือการนำหัวอ่านที่ได้จากกระบวนการผลิตหัวเขียนอ่านข้อมูล มาประกอบเข้ากับแกนของหัวอ่านเขียน หรือ flexer ที่มีแผงวงจรไฟฟ้าอยู่เรียบร้อยแล้ว ในกระบวนการนี้จะมีขั้นตอนการทดสอบค่าที่สำคัญซึ่งเกี่ยวกับทางไฟฟ้า เป็นการทดสอบการทำงานของหัวอ่านเขียนทั้งก่อน และหลังจากต่อวงจรแล้วว่ามีประสิทธิภาพการอ่านเขียนเป็นเท่าไร
4. กระบวนการผลิต Head Stack Assembly หรือ HSA คือ การนำหัวอ่านเขียนที่ประกอบแกนหัวอ่านเขียนแล้วจากกระบวนการผลิต Head Gimbal Assembly มารวมกัน โดยใน 1 HSA อาจมีได้หลายแกนและหลายหัวอ่านเขียนขึ้นอยู่กับการออกแบบขนาดความจุของฮาร์ดดิสก์ ในขั้นตอนนี้มีการทดสอบคล้ายๆ กันกับกระบวนการผลิต HGA คือมีการทดสอบค่าทางไฟฟ้าเช่นกัน เพื่อดูประสิทธิภาพการอ่านเขียนหลังจากทำการประกอบรวมกันหลายแกนหัวอ่านเขียน
5. กระบวนการผลิต Hard Disc Drive หรือ HDD คือการนำ HSA มาประกอบกับตัว Drive และทดสอบทั้งชุด กระบวนการประกอบเป็น Drive เป็นกระบวนการผลิตที่มีขั้นตอนจำนวนมาก ตั้งแต่ นำ HSA และแผ่นจานหมุน มาประกอบเข้าชุดกล่องวงจรไฟฟ้าภายในห้องสะอาด เพื่อป้องกันฝุ่นละอองเข้าไปทำให้หัวอ่านเขียนเสียหาย ซึ่งแผ่นจานหมุนและชุดกล่องวงจรไฟฟ้ามีการสั่งทำจากโรงงานภายนอกซึ่งจะมีเกณฑ์การวัดคุณภาพของสินค้าที่นำส่งเข้ามาทำการผลิตหลังจากประกอบเป็นฮาร์ดดิสก์สำเร็จรูปแล้วจะทำการทดสอบหลังทำการประกอบเสร็จแล้วว่าเมื่อทำการประกอบครบชุดแล้วฮาร์ดดิสก์นั้นให้ประสิทธิภาพการอ่านเขียนได้เท่าไหร่ซึ่งส่วนใหญ่ใช้เครื่องจักรอัตโนมัติในการทดสอบมากกว่าใช้แรงงานคนหลังจากนั้นจะทำการคัดเกรดของผลิตภัณฑ์พร้อมบรรจุเพื่อนำส่ง

ออกนอกประเทศต่อไปเป็นการเสร็จสิ้นกระบวนการผลิต ดังแสดงลำดับขั้นตอนการผลิตในภาพที่ 2-1



ภาพที่ 2-1 ลำดับขั้นตอนการผลิตฮาร์ดดิสก์

จะเห็นได้ว่าแต่ละขั้นตอนของกระบวนการผลิตจะต้องมีการทดสอบคุณภาพของชิ้นงานเสมอเพื่อตรวจสอบว่าชิ้นงานที่ผลิตมานั้นได้คุณภาพตามมาตรฐานที่กำหนดไว้หรือไม่ก่อนทำการนำส่งต่อในกระบวนการต่อไป

งานวิจัยนี้สนใจเฉพาะในกระบวนการผลิต HGA ซึ่งเป็นกระบวนการประกอบหัวอ่านเขียนกับแกนหัวอ่านเขียนซึ่งในกระบวนการนี้จะมีการทดสอบค่าต่างๆ มากมายโดยเฉพาะค่าทางไฟฟ้าที่ต้องทำการวัดก่อนและหลังจากทำการประกอบหัวอ่านเขียนเข้ากับแกนหัวอ่านเขียน และค่าทางกลศาสตร์ ที่เกิดจากการวัดแกนหัวอ่านเขียน ซึ่งค่าเหล่านี้ได้มาจากขั้นตอนการผลิต เครื่องจักร และคุณสมบัติของตัวงาน การวัดแต่ละครั้งเราเรียกชื่อค่าที่วัดได้นั้นเป็นพารามิเตอร์ ซึ่งในการผลิตมีพารามิเตอร์จำนวนมากที่ทำการทดสอบ แต่งานวิจัยนี้สนใจเฉพาะ

กลุ่มของพารามิเตอร์ที่มีผลต่อคุณภาพของชิ้นงาน งานวิจัยนี้จึงแบ่งพารามิเตอร์ออกเป็น 3 ประเภท คือ

1. พารามิเตอร์ที่ไม่สามารถทำการปรับค่าได้ ได้แก่ พารามิเตอร์ที่เกิดจากการวัดค่าทางกายภาพ เช่น ความกว้าง ความยาว ความลึก ความห่าง ซึ่งค่าเหล่านี้ไม่สามารถทำการเปลี่ยนแปลงได้ เพราะเป็นค่าที่มาจากตัวชิ้นงาน
2. พารามิเตอร์ที่สามารถทำการปรับค่าได้ ได้แก่ พารามิเตอร์ที่สามารถทำการเปลี่ยนแปลงได้โดยทันที เช่น การหมุนปรับค่า การเปลี่ยนสูตร การลด หรือ เพิ่มตัวแปร ขนาดเส้นผ่านศูนย์กลางของสารที่หยอดโดยเครื่องจักร ซึ่งค่าพารามิเตอร์นี้สามารถทำการปรับค่าได้ทันทีเมื่อมีการตัดสินใจให้ทำการเปลี่ยนแปลง
3. พารามิเตอร์ที่มีความสัมพันธ์กับพารามิเตอร์ที่สามารถทำการปรับค่าได้ ได้แก่ พารามิเตอร์ที่ได้รับผลกระทบเมื่อพารามิเตอร์ที่สามารถทำการปรับค่าได้เกิดการเปลี่ยนแปลงไป

งานวิจัยนี้มีจุดมุ่งหมายที่จะสร้างแบบจำลองในการพัฒนาคุณภาพของชิ้นงานให้ดีขึ้นโดยมุ่งเน้นไปที่พารามิเตอร์ที่สามารถทำการปรับค่าได้และพารามิเตอร์ที่มีความสัมพันธ์กับพารามิเตอร์ที่สามารถทำการปรับค่าได้เพราะสามารถทำการแก้ไขได้ทันทีในกระบวนการที่เกิดปัญหาและไม่กระทบกับกระบวนการอื่นๆ โดยชิ้นงานที่จะนำมาสร้างแบบจำลองนั้นต้องมีการออกแบบขั้นตอนการผลิตที่แน่นอนและใช้ประจำอยู่แล้ว คือมีขั้นตอนการผลิตซ้ำเดิมในทุกๆวัน เพื่อให้ข้อมูลปัจจุบันที่นำมาทดลองนั้น สามารถนำไปใช้ปรับปรุงกระบวนการผลิตสินค้าเดิมๆ ได้ในอนาคต

2.1.2 **ต้นไม้ตัดสินใจ** (Decision Tree)

ต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่องที่นิยมใช้มากที่สุดแบบหนึ่ง [2] โดยการจำแนก (classification) ข้อมูลออกเป็นคลาส (class) ต่างๆ โดยใช้คุณสมบัติ (attribute) ของข้อมูลในการจำแนกว่าคุณสมบัติใดของข้อมูลที่เป็นตัวกำหนดการจำแนกและคุณสมบัติแต่ละตัวของข้อมูลมีการวัดความสำคัญอย่างไร

ต้นไม้ตัดสินใจ ประกอบไปด้วย [3]

1) โหนดภายใน (internal node) คือ คุณสมบัติต่างๆ ของข้อมูล ใช้ในการตัดสินใจว่าข้อมูลจะไปอยู่ในกรณีไหน โดยโหนดภายในที่เป็นโหนดเริ่มต้น เรียกว่า โหนดราก
2) กิ่ง (branch, link) เป็นค่าคุณสมบัติหรือเงื่อนไขของคุณสมบัติในโหนดที่ใช้ในการจำแนกข้อมูล ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าคุณสมบัติของโหนดภายในนั้น
3) โหนดใบ (leaf node) คือคลาสต่างๆ ซึ่งเป็นผลลัพธ์ในการจำแนกข้อมูล

ลักษณะการเรียนรู้ของต้นไม้ตัดสินใจ

- ผลการเรียนรู้แสดงอยู่ในรูปที่เข้าใจง่าย ซึ่งง่ายต่อการวิเคราะห์คุณสมบัติที่มีผลต่อการจำแนกคลาสต่างๆ
- แต่ละเส้นทางจากโหนดรากถึงโหนดใบสามารถแสดงให้อยู่ในรูปกฎ IF-THEN ได้
- มีความทนทานต่อข้อมูลรบกวน (noisy data)
- การเรียนรู้มีความรวดเร็วเมื่อเทียบกับอัลกอริทึมสำหรับจำแนกชนิดอื่น
- เหมาะแก่การนำไปใช้ในการวิเคราะห์งานทางด้านธุรกิจ ความเสี่ยงของลูกหนี้

ซึ่งในงานวิจัยนี้ต้องการนำการเรียนรู้ของต้นไม้ตัดสินใจมาใช้ในการวิเคราะห์หาพารามิเตอร์ที่มีผลต่อคุณภาพชิ้นงาน โดยจำแนกคุณสมบัติวัตถุดิบที่นำเข้ามาและก่อให้เกิดคลาสต่างๆ โดยสามารถนำเสนอข้อมูลในรูปของกฎได้ เพื่อใช้ในการหาความสัมพันธ์และคุมพารามิเตอร์ในการผลิตได้

วิธีการเรียนรู้ของต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจจะเป็นการค้นหาจากบนลงล่างแบบตะกราม (top-down greedy search) โดยเริ่มจากเลือกคุณสมบัติที่ดีที่สุดในการแยกคลาสมาเป็นโหนดราก เมื่อข้อมูลผ่านการแบ่งที่โหนดรากแล้ว ก็จะหาคุณสมบัติที่ดีที่สุดของข้อมูลที่ถูกแบ่งออกแล้วมาสร้างโหนด

ลูกของรากนั้นต่อไป และจะวนสร้างโหนดลูกและต้นไม้ย่อยของแต่ละกิ่งไปเรื่อยๆ จนกว่าข้อมูลที่ผ่านการแบ่งแยกนั้นจะจัดอยู่ในคลาสเดียวกัน

การนิยามค่าความดีของคุณสมบัตินั้นแตกต่างกันไปตามอัลกอริทึม ในงานวิจัยนี้เลือกใช้ค่ามาตรฐานเกน (Gain criterion) ของอัลกอริทึม C4.5 ซึ่งเป็นอัลกอริทึมที่แพร่หลายที่สุด

อัลกอริทึม C4.5

การสร้างต้นไม้ตัดสินใจแบบ C4.5 คล้ายกับอัลกอริทึม ID3 แต่มีการพัฒนาเพิ่มเติม กล่าวคือ จะใช้ค่ามาตรฐานเกนในการตัดสินใจเลือกคุณสมบัติที่ใช้เป็นรากหรือโหนดในต้นไม้ โดยการคำนวณค่าเกณฑ์จะเลือกคุณสมบัติที่มีค่าเกนสูงที่สุดมาเป็นรากหรือโหนด ค่าเกนนี้คำนวณได้โดยใช้ความรู้จากทฤษฎีสารสนเทศ คือ ค่าสารสนเทศของข้อมูลขึ้นอยู่กับค่าความน่าจะเป็นของข้อมูล ซึ่งสามารถวัดได้ในรูปแบบของบิต (bits) ตามสูตร

ค่าสารสนเทศของ M หรือ ค่าเอนโทรปีของ M เขียนแทนด้วย I(M) จากสูตร

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

โดย ถ้าให้ชุดของข้อมูล M ประกอบด้วยค่าที่เป็นไปได้ คือ $\{m_1, m_2, \ldots, m_n\}$ ความน่าจะเป็นที่จะเกิดค่า m_i เท่ากับ $P(m_i)$

จากสูตรแสดงให้เห็นว่าค่าสารสนเทศที่น้อยหมายถึงข้อมูลชุดนั้นมีความแตกต่างกันน้อย แต่ถ้าค่าสารสนเทศมากหมายถึงข้อมูลชุดนั้นมีความแตกต่างกันมาก หรือ ประกอบด้วยตัวอย่างหลายพวกที่มีจำนวนใกล้เคียงกัน

ในการเลือกคุณสมบัติที่จะมาเป็นโหนดรากจะใช้ค่ามาตรฐานเกน นั้นคำนวณจากค่าสารสนเทศทั้งหมดของชุดข้อมูลนั้นลบด้วยค่าสารสนเทศหลังจากเลือกคุณสมบัติใดคุณสมบัติหนึ่งเป็นราก

ค่าสารสนเทศหลังจากแบ่งคุณสมบัติคำนวณได้จากค่าผลรวมของผลคูณระหว่างค่าสารสนเทศของแต่ละโหนดกับอัตราส่วนของตัวอย่างในแต่ละกิ่งต่อตัวอย่างทั้งหมดที่โหนดนั้น หรือความน่าจะเป็นไปได้ของแต่ละคุณสมบัติ

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

โดยที่ T คือ ข้อมูลเรียนรู้ และ X คือคุณสมบัติที่เป็นโหนด ซึ่งมีค่าเป็นไปได้ n ค่า โหนดที่แบ่งตามคุณสมบัติ X จะสามารถแบ่งข้อมูล T ออกมาเป็นกิ่ง $\{t_1,t_2,t_3,\ldots,t_n\}$ ตามค่าโหนดที่เป็นไปได้

ค่ามาตรฐานของคุณสมบัติ X สามารถคำนวณได้จากการลบค่าสารสนเทศทั้งหมดที่โหนดนี้กับค่าสารสนเทศที่ได้หลังจากแบ่งด้วยคุณสมบัติ X

$$Gain(X) = I(T) - I_x(T)$$

ในอัลกอริทึม C4.5 ได้เพิ่มการใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio criterion) ในการตัดสินใจเลือกคุณสมบัติที่จะใช้เป็นรากหรือโหนดอีกอย่างหนึ่ง เนื่องจากค่าเกนมาตรฐานจะมีค่าไบแอส(Bias) อย่างมากกับข้อมูลที่ประกอบด้วยคุณสมบัติที่มีค่าเป็นไปได้จำนวนมากๆ เช่น ตัวอย่างที่ไม่ซ้ำกัน หรือ ตัวอย่างที่ประกอบด้วยข้อมูลคลาสเดียว การแก้ไขความอคติของค่ามาตรฐานเกนสามารถทำได้โดยการปรับค่ามาตรฐานเกนให้ถูกต้อง โดยใช้ค่าสารสนเทศของการแบ่งแยก (split information) ของคุณสมบัติแต่ละตัว ถ้าให้ T เป็นชุดข้อมูลแบ่งตามคุณสมบัติ X และได้ชุดของตัวอย่างย่อยในแต่ละกิ่ง n ชุด ตามค่าที่เป็นไปได้ในคุณสมบัติ X ทำการคำนวณค่าสารสนเทศของการแบ่งแยกได้ ดังนี้

ค่าสารสนเทศของการแบ่งแยก $= -\sum_{i-1}^{n} \frac{|\mathrm{t_i}|}{|\mathrm{T}|} \log_2 \frac{|t_i|}{|T|}$

เมื่อนำค่าสารสนเทศของการแบ่งแยกไปหารค่ามาตรฐานเกนจะได้ค่ามาตรฐานอัตราส่วนเกน ซึ่งช่วยแก้ไขความอคติของค่ามาตรฐานเกนได้ โดยทำให้ค่ามาตรฐานอัตราส่วนเกนในการแบ่งด้วยคุณสมบัติที่มีการกระจายสูงถูกปรับลดลง ดังนั้นค่ามาตรฐานอัตราส่วนเกนในคุณสมบัติของตัวอย่างที่มีการกระจายตัวของข้อมูลสูงดังที่กล่าวมาแล้วจะไม่มีค่าสูงที่สุดเสมอ

การตัดเล็มต้นไม้ตัดสินใจ

ในการสร้างต้นไม้ตัดสินใจจะแบ่งข้อมูลจนกระทั้งได้ข้อมูลที่เป็นคลาสเดียวกันหมด แต่จะเฉพาะเจาะจงกับข้อมูลที่ใช้ในการเรียนรู้เท่านั้น หรือ เรียกว่า ภาวะเหมาะเกินไป ยิ่งต้นไม้ตัดสินใจที่มีขนาดใหญ่ขึ้นจะให้ข้อมูลความถูกต้องบนข้อมูลสอนมากขึ้น แต่เมื่อนำไปใช้งานจริงกลับทำให้ความความถูกต้องลดลง

ดังนั้นต้นไม้ควรทำการตัดเล็มเพื่อให้ได้ต้นไม้ขนาดเล็กลง และ ลดความเฉพาะเจาะจงกับข้อมูลที่ใช้สอน

ในอัลกอริทึม C4.5 จะตัดเล็มโดยใช้ค่าความผิดพลาด (error-based pruning) คือมีการรวมต้นไม้ย่อยเข้าด้วยกันเป็นโหนดเดียวกันก็ต่อเมื่อหลังจากรวมแล้วไม่ทำให้ค่าความผิดพลาดเพิ่มขึ้น โดยค่าความผิดพลาดที่ใช้ทดสอบกับข้อมูลที่ไม่เคยเห็นจะใช้ค่าจำกัดบนของการกระจายแบบไบโนเมียล (binomial distribution) ที่ระดับความเป็นอิสระ CF (confidence level)

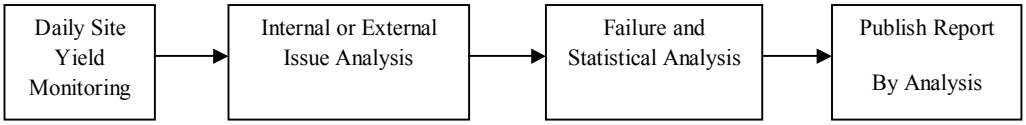
ถ้ามีข้อมูล N ตัวที่โหนด และมีข้อมูล E ตัวเป็นข้อมูลที่มีคลาสไม่ถูกต้อง ค่าความผิดพลาดในโหนดนี้เขียนได้เป็น $U_{cf}(E,N)$ โดยมีข้อกำหนดที่ว่าขนาดของตัวอย่างสอนเท่ากับขนาดตัวอย่างของข้อมูลที่ไม่เคยเห็น ซึ่งสามารถคาดได้ว่าจะมีจำนวนข้อมูลที่จำแนกผิดพลาดเท่ากับ $N \times U_{cf}(E,N)$ ตัว ซึ่งถ้าคำนวณจำนวนข้อมูลที่คาดว่าจะจำแนกผิดพลาดของแต่ละกิ่งรวมกันแล้วมากกว่าจำนวนข้อมูลที่คาดว่าจะจำแนกผิดพลาดของโหนดที่แตกกิ่งนั้น ก็จะตัดโหนดที่เป็นลูกในทุกกิ่งของโหนดที่แตกกิ่งนั้นออกให้หมดจนเหลือเฉพาะโหนดที่แตกกิ่งนั้นไว้โหนดเดียว

2.1.3 Weka

โปรแกรม Weka (Waikato Environment for Knowledge Analysis) [4] พัฒนาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นโปรแกรมที่รวบรวมอัลกอริทึมของการเรียนรู้ด้วยเครื่องในการทำเหมืองข้อมูล พัฒนาด้วยภาษาจาวาเป็นซอฟต์แวร์สำเร็จประเภทฟรีแวร์ ภายใต้การควบคุมของ GNU General Public License ซึ่งประกันการใช้ซอฟต์แวร์อย่างเสรีภาพWeka มีเครื่องมือที่สนับสนุนการทำเหมืองข้อมูลซึ่งประกอบไปด้วย การเตรียมข้อมูล (data pre-processing), การจำแนกประเภท (classification), การถดถอย (regression), การจัดกลุ่ม (clustering), กฎความสัมพันธ์ (association rules) และการสร้างภาพ (visualization) ซึ่งเหมาะสมที่จะใช้เป็นเครื่องมือที่จะช่วยพัฒนารูปแบบการเรียนรู้ด้วยเครื่องที่จะใช้ในงานวิจัยนี้

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

อัลกอริทึมที่ใช้จำแนกความดีของคุณลักษณะในต้นไม้ตัดสินใจมีมากมายได้แก่ ID3 [4], C4.5 [5], CART [6] และ CHAID [7] มีงานวิจัยที่เกี่ยวกับ C4.5 เพื่อใช้ในการหารูปแบบของงานเสียในการผลิต HGA ซึ่งเป็นส่วนประกอบของฮาร์ดดิสก์ งานวิจัย [8] ได้นำเสนองานวิจัยที่เกี่ยวกับการออกแบบอัลกอริทึมสำหรับวิเคราะห์ความผิดพลาดด้วยเทคนิคการสร้างต้นไม้ตัดสินใจเพื่อวิเคราะห์หาสาเหตุที่ทำให้ผลผลิตตกต่ำของกระบวนการผลิตฮาร์ดดิสก์ที่กระบวนการผลิต HGA ซึ่งเกิดจากผลการผลิตก่อนหน้าของกระบวนการผลิตที่ Wafer และ Slider โดยผลที่ได้สามารถระบุคุณสมบัติที่เป็นสาเหตุในรูปแบบของ เครื่องจักร วัตถุดิบ หรือวิธีการ ซึ่งในปัจจุบันวิธีการค้นหาปัญหาในโรงงาน ยังใช้บุคคลากรในการวิเคราะห์หาสาเหตุด้วยวิธีการวิเคราะห์เชิงสถิติ ซึ่งเสียเวลามาก ดังแสดงขั้นตอนการทำงานของบุคลากรดัง*ภาพที่2-2*

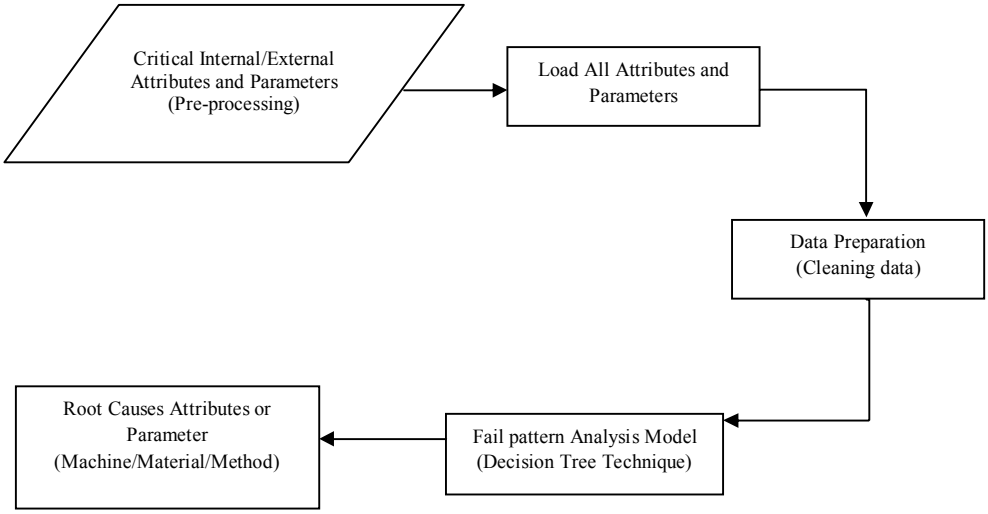


ภาพที่ 2-2 ผังงานของการวิเคราะห์ปัญหา

งานวิจัยนี้จึงช่วยพัฒนาการออกแบบอัลกอริทึมสำหรับวิเคราะห์ความผิดพลาดเพื่อช่วยลดต้นทุนและเวลาการทำงานลง โดยนำเสนอการใช้เทคนิคต้นไม้ตัดสินใจ ซึ่งต้นไม้

ตัดสินใจจะช่วยในการจัดเตรียมข้อมูลให้อยู่ในรูปแบบโครงสร้างต้นไม้เพื่อนำไปวิเคราะห์ กำหนดให้ โหนดภายในแทนคุณสมบัติและพารามิเตอร์ กำหนดให้คลาสที่ถูกจำแนกแทนด้วย โหนดใบ

งานวิจัยนี้เลือกใช้อัลกอริทึม C4.5 หลังจากทำการเปรียบเทียบอัลกอริทึมของสร้างต้นไม้ตัดสินใจที่เหมาะสมอยู่ 4 แบบ คือ ID3, C4.5, CARTและCHAID เนื่องจากอัลกอริทึม C4.5 ถูกพัฒนามาจากอัลกอริทึม ID3 ให้เพิ่มประสิทธิภาพในการคำนวณ จัดการได้ดีกับค่าต่อเนื่อง และ คุณสมบัติที่เป็นค่าว่าง ดังนั้น อัลกอริทึม C4.5 จึงดีกว่า อัลกอริทึม ID3 สำหรับอัลกอริทึม CART นั้น ในโหนดใบต้องเป็นประเภทค่าต่อเนื่องเท่านั้น ไม่ใช่ประเภทค่าที่แยกกันได้ ซึ่งข้อมูลงานวิจัยนี้ต้องการโหนดใบที่เป็นประเภทค่าที่แยกกันได้ ดังนั้น อัลกอริทึม CART จึงไม่เหมาะกับข้อมูลการวิจัยนี้ ในขณะที่โหนดใบ ของอัลกอริทึม C4.5 สามารถเป็นได้ทั้ง ประเภทค่าที่แยกกันได้ หรือ ประเภทค่าต่อเนื่อง อีกทั้งอัลกอริทึม CART มีโครงสร้างแบบไบนารีเท่านั้นทำให้แบ่งคลาสได้ทีละ 2 ในขณะที่อัลกอริทึม C4.5 สามารถแบ่งกิ่งได้มากกว่า 2 กิ่ง ไม่จำกัดว่าต้องเป็นโครงสร้างแบบไบนารี เท่านั้น ส่วนอัลกอริทึม CHAID ไม่สนับสนุนการตัดเล็มต้นไม้ตัดสินใจ จึงไม่เหมาะที่จะนำมาใช้กับข้อมูลขนาดใหญ่ของโรงงาน แต่อัลกอริทึม C4.5 มีความสามารถในการตัดเล็มต้นไม้ตัดสินใจ แก้ปัญหาข้อมูล โอเวอร์ฟิตติ้ง ได้ ดังนั้นงานวิจัยนี้จึงเลือกใช้อัลกอริทึม C4.5 ในการสร้าง model ซึ่งมีขั้นตอนการทำงานดัง*ภาพที่ 2-3*



ภาพที่ 2-3 ผังงานของโมเดลการวิเคราะห์วิเคราะห์ความผิดพลาด

ขั้นตอนที่ 1 ข้อมูลที่นำเข้ามาต้องมีการเลือกคุณสมบัติและพารามิเตอร์ที่สำคัญไว้ก่อนการนำข้อมูลเข้า โปรแกรมในขั้นตอนที่ 2

ขั้นตอนที่ 3 เพื่อทำความสะอาดข้อมูล โดยแบ่งออกเป็น 3 ขั้นตอน คือ

- Indices replacement คือการเปลี่ยนค่าข้อมูลที่เป็นชื่อให้แทนด้วยตัวเลข
- Column displacement ใช้กำจัดคอลัมน์ที่เป็นค่าว่าง
- Missing values treatment จัดการค่าที่หายไปด้วยการใช้ the corrected gain ratio method
- Data correlation ใช้ตระเตรียมข้อมูลในโมเดลในการหาค่าความสัมพันธ์ของทุกคุณสมบัติและพารามิเตอร์ โดยใช้ The sampled correlation coefficient

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)S_x S_y}$$

กำหนดให้ X เป็นเซตที่ประกอบไปด้วยทุกพารามิเตอร์และทุกคุณสมบัติ

Y เป็นเซตของพารามิเตอร์ที่สำคัญ

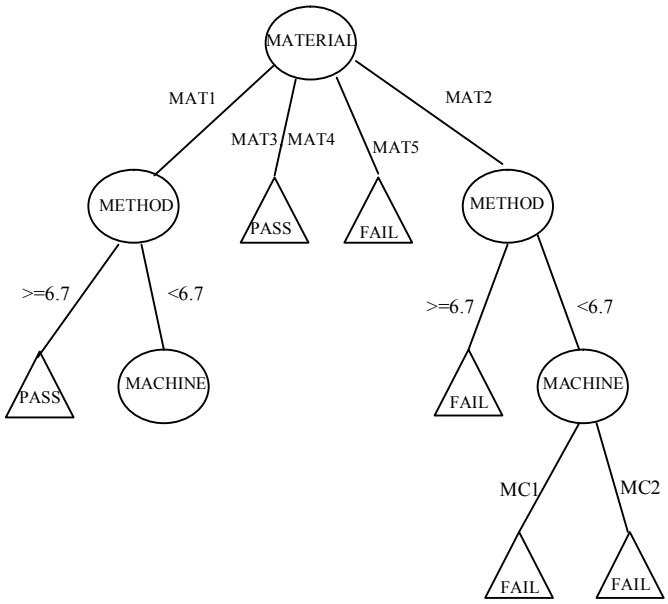
n เป็นจำนวนของข้อมูล

$\bar{X}$ และ $\bar{Y}$ เป็นค่าเฉลี่ยของค่า X และ Y

S_{x} และ S_{y} เป็นค่าการกระจายตัวของข้อมูลของค่า X และ Y

ขั้นตอนที่ 4 การสร้างโมเดลด้วยต้นไม้ตัดสินใจอัลกอริทึม C4.5 ในการหาค่าเกนที่สูงที่สุด เพื่อสร้างโหนดราก และโหนดอื่นๆ

ขั้นตอนที่ 5 จากโมเดลต้นไม้ตัดสินใจที่ได้จะสามารถระบุต้นเหตุของปัญหาได้ โดยในงานวิจัยนี้ได้ทำการแบ่งคุณสมบัติออกเป็น 3 ประเภท คือ เครื่องจักร, วัตถุดิบ และวิธีการ ดังแสดงใน*ภาพที่ 2-4*



ภาพที่ 2-4 ตัวอย่างต้นไม้ตัดสินใจที่ได้

งานวิจัยนี้สามารถสรุปผลการทดลองจากการทดลองกับ 20 คุณสมบัติและจำนวนข้อมูล 1000 ข้อมูลได้ว่าสามารถทำนายว่าผลิตภัณฑ์ใหม่นั้นจะเป็นผลิตภัณฑ์ที่ดีหรือเสียภายใต้การตั้งค่าตามที่ระบุของคุณสมบัติ จากรูปแบบที่สร้างขึ้น

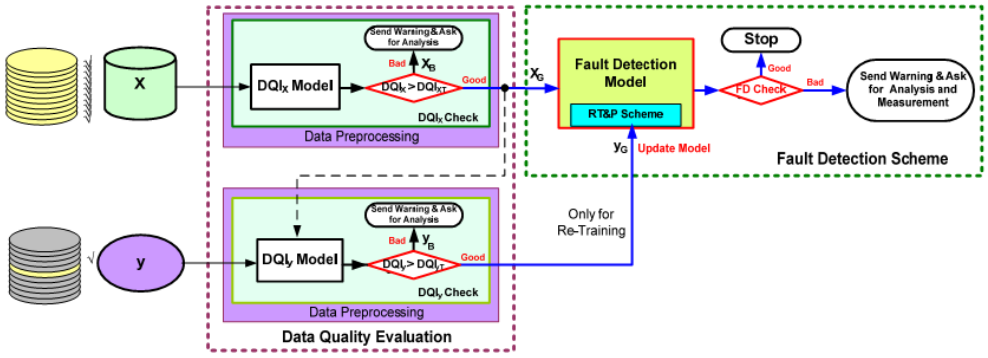
งานวิจัย [9] เกี่ยวข้องกับการตรวจจับของเสียในการผลิตสารกึ่งตัวนำโดยใช้ CART ในการสร้างแบบจำลองเพื่อระบุความสัมพันธ์ระหว่างพารามิเตอร์กับกระบวนการผลิต โดยประยุกต์ใช้หลักการต้นทุนต่ำสุดสำหรับตัดเล็มต้นไม้เพื่อหลีกเลี่ยงสภาวะข้อมูลเหมาะเกินไป งานวิจัยนี้ใช้ในการตรวจจับงานเสียของโรงงาน current semiconductor และ TFT-LCD ด้วย FDS (Fault Detection Scheme) ซึ่งเป็นประโยชน์ในการจัดคลาสและนำต้นไม้ถดถอยไปปรับปรุงโมเดลเพื่อระบุความสัมพันธ์ระหว่างพารามิเตอร์ของกระบวนการและผลิตภัณฑ์ที่ไม่ผ่านเกณฑ์คุณภาพ โดยที่โมเดลนี้สามารถใช้งานบนเวลาจริงเพื่อตรวจจับของเสีย

แม้ว่าปัจจุบันกระบวนการทำงานของโรงงานมีระบบตรวจจับ a fault-detection-and-classification (FDC) server อยู่แล้ว แต่ยังพบความผิดพลาดในการตรวจจับอยู่ เช่น การ

ตรวจจับพารามิเตอร์ในกระบวนการผลิตพบว่าอยู่ในเกณฑ์ที่ตั้งไว้ แต่ ผลของคุณภาพผลิตภัณฑ์รวมกลับไม่อยู่ในเกณฑ์ที่ตั้งไว้ ทั้งนี้เนื่องจากการวัดคุณภาพผลิตภัณฑ์ใช้การสุ่มเลือกขึ้นมาตามโอกาส ซึ่งไม่เหมือนกับพารามิเตอร์ที่ถูกเก็บข้อมูลมาจากเครื่องมือในการผลิต

งานวิจัยนี้จึงเสนอแนวทางแก้ปัญหานี้ด้วยวิธี FDS (Fault Detection Scheme) เริ่มต้นด้วยการสร้างโมเดล FD model (Fault Detection model) ซึ่งใช้เทคนิคเหมืองข้อมูลอัลกอริทึมที่เรียกว่า “classification and regression tree (CART)” ช่วยในการสร้างโครงสร้างของโมเดลเพื่อระบุความสัมพันธ์ระหว่างพารามิเตอร์ของกระบวนการและผลิตภัณฑ์ที่อยู่ในเกณฑ์ (OOS product)

ขั้นตอนการสร้างโมเดล FD เป็นดังแสดงใน*ภาพที่2-5*



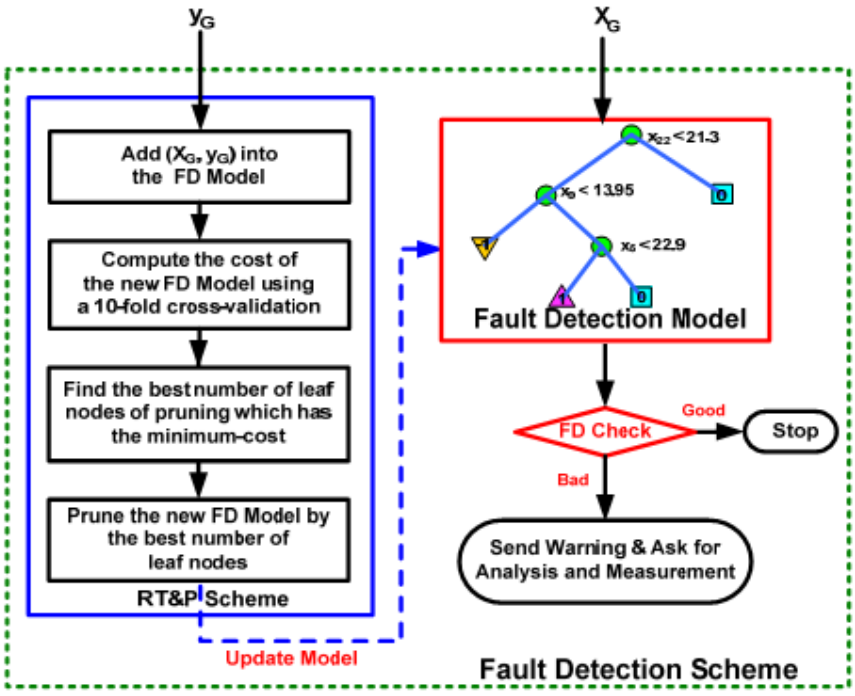
ภาพที่ 2-5 แผนผังการทำงานโดยรวมของ FD โมเดล

ขั้นตอนที่ 1 Data Collection ขั้นตอนนี้จะทำการเก็บรวบรวมข้อมูลของกระบวนการจนสามารถนำไปสร้าง FD model ได้ ซึ่งในกระบวนการนี้จะต้องทำการนำข้อมูลที่ผิดปกติออกไปก่อน

ขั้นตอนที่ 2 Data Quality Evaluation หลังจากทำการเก็บข้อมูลได้มากพอจะสร้าง FD model แล้วจะทำการสร้างเพื่อใช้วัดคุณภาพขั้นแรกก่อนใช้ FD model โดยโมเดลเริ่มต้น DQI_x ถูกสร้างขึ้นก่อน แล้วจึงนำโมเดล DQI_x มาสร้าง โมเดล DQI_y

โมเดล DQI_x ที่ใช้ในการวิเคราะห์งานถูกสร้างขึ้นจาก principal component analysis (PCA) ซึ่งถูกประยุกต์ใช้ในการวิเคราะห์ข้อมูลกระบวนการทั้งหมด จากนั้นจึงใช้ Euclidean distance จัดให้อยู่ในรูปของคุณภาพข้อมูลดัชนี หรือ DQI_x สำหรับ โมเดล DQI_y ใช้ทฤษฎี adaptive resonance theory 2 (ART2) และ normalized variability (NV) ในการสร้างดัชนี

ขั้นตอนที่ 3 Fault Detection Model เป็นการสร้างโมเดลขึ้นมาโดยใช้อัลกอริทึม CART ในการสร้าง โดยโมเดลนี้สามารถใช้ในการตรวจจับงานบนเวลาจริง ให้ผลแม่นยำและหลีกเลี่ยงกรณี การเกิดการเตือนที่ผิด (FA) และ การผิดพลาดในการตรวจจับ (MD) ซึ่งในโรงงาน MD ถือเป็นกรณีที่ร้ายแรงกว่า FA การตัดเล็มต้นไม้ถูกออกแบบเป็นโมเดล FD ด้วยหลักการต้นทุนน้อยที่สุดในการตัดเล็มต้นไม้ การตั้งต้นทุนกำหนดให้ถ้าเกิดกรณี MD ต้นทุนเป็น 2 FA ต้นทุนเป็น 1 และ ตรวจจับถูกต้อง (CD) ต้นทุนเป็น 0 ซึ่งขั้นตอนลำดับการสร้างโมเดลเป็นดัง*ภาพที่ 2-6*



ภาพที่ 2-6 **แผนผังการสร้าง** FD model

ซึ่งผลจากการทดลองพบว่า หลังจากทำการเรียนรู้อีกครั้งด้วยการตัดเล็มต้นไม้ พบว่า งาน CD เพิ่มมากขึ้น ในขณะที่ FAและ MD ลดน้อยลง มีความถูกต้องแม่นยำมากขึ้น ดังแสดงใน*ตารางที่ 2-1*

TABLE I
FDS EXPERIMENTAL RESULTS

196 testing samples	Without Pruning					With Pruning				
	Case	CD	FA	MD	Accuracy	Case	CD	FA	MD	Accuracy
Free-Running Mode	#1	174	21	1	0.888	#2	183	13	0	0.933
Re-Training Mode	#3	193	2	1	0.985	#4	195	1	0	0.995

ตารางที่ 2-1 ตารางการทดลอง FDS EXPERIMENTAL RESULTS

งานวิจัยต่อมา [10] ได้นำเสนองานวิจัยการออกแบบเครื่องมือในการวิเคราะห์พัฒนาคุณภาพการผลิตของอุตสาหกรรมฮาร์ดดิสก์ โดยทำการหาพารามิเตอร์และคุณลักษณะที่มีผลกระทบต่อคุณภาพของผลิตภัณฑ์ด้วยต้นไม้ตัดสินใจและทำการพัฒนาคุณภาพด้วยการปรับค่าเฉลี่ยของพารามิเตอร์เพื่อให้ได้คุณภาพของผลิตภัณฑ์ที่สูงขึ้น ประกอบไปด้วยอัลกอริทึมที่หลากหลายในการวิเคราะห์ข้อมูลอัตโนมัติ แบ่งออกเป็น 2 โมดูล

โมดูลแรก ใช้เทคนิค Sensitivity ในการค้นหาตัวแปรที่สำคัญ และข้อจำกัดของผลิตภัณฑ์ (Product Specification) ที่มีผลต่อปริมาณผลผลิต

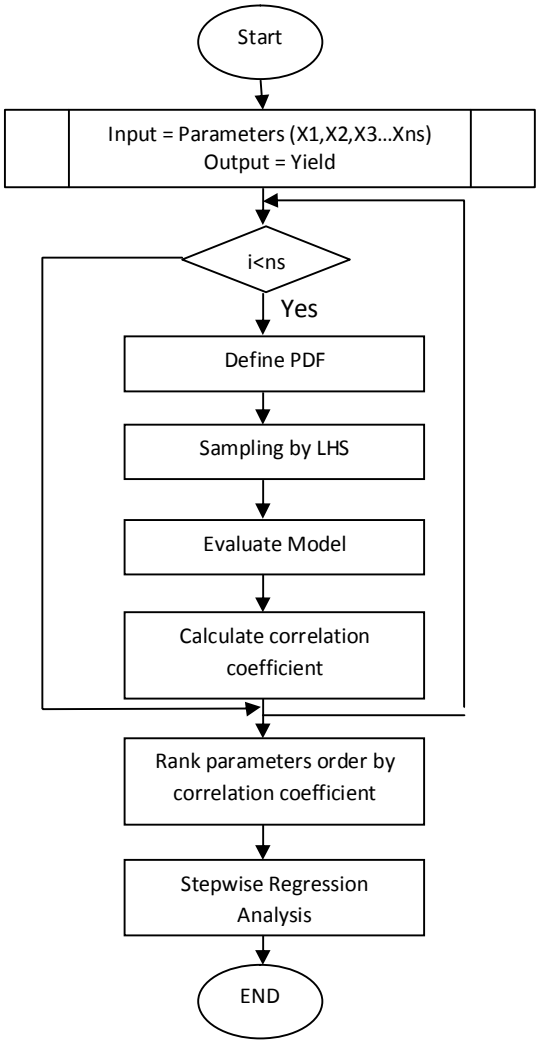
โมดูลที่ 2 จะเรียนรู้และทำการพัฒนาวิธีเพื่อใช้ในอุตสาหกรรมในการพัฒนาปริมาณผลผลิตด้วยการ การขยับค่าเฉลี่ย (mean shift), การลดค่าความแปรปรวน (variance reduction) และการกระชับข้อจำกัด (Spec Tightening)

งานวิจัยนี้ใช้เทคนิคการวิเคราะห์ข้อมูลด้วย Sensitivity ในการหาเซตของตัวแปรและคุณสมบัติที่สำคัญ ที่มีผลต่อ ปริมาณผลผลิตซึ่งหลังจากค้นพบตัวแปรที่สำคัญแล้วจะทำการวิเคราะห์ต่อเพื่อหาที่มาของปัญหาด้วยอัลกอริทึม Decision Tree

การวิเคราะห์ด้วยเทคนิค Sensitivity แบ่งออกดังนี้

1) การวิเคราะห์ตัวแปรที่สำคัญ

งานวิจัยนี้สนใจเฉพาะตัวแปรที่มีผลกระทบต่อปริมาณผลผลิตของกระบวนการผลิต ซึ่งถือเป็นปัญหาคลอบคลุมรวม ใช้วิธี Sampling-based sensitivity analysis โดยมีการทำงานตาม flow ดังแสดงใน*ภาพที่ 2-7*



ภาพที่ 2-7 Flow **การทำงาน ด้วยวิธี** Sampling-based sensitivity

จากแผนผัง เริ่มต้นข้อมูลที่นำเข้ามาคือตัวแปรต่างๆ ที่น่าจะมีผลต่อปริมาณผลผลิต จากนั้นใช้อัลกอริทึม probability density-function กับตัวแปรที่เข้ามาและสุ่มการ

กระจายตัว (Sampling Distribution) ด้วย the Latin Hypercube Sampling (LHS) ซึ่งเป็นวิธีที่คลอบคลุมขอบเขตได้ทั้งหมด จากนั้นทำการจับคู่ระหว่างค่าปริมาณผลผลิตกับ กลุ่มเซตของข้อมูลที่นำเข้ามาทำการวิเคราะห์ Sensitivity ด้วยค่า correlation coefficients ด้วยการวัด Linear ระหว่าง ตัวแปรที่เข้ามากับปริมาณผลผลิต ตามสมการ

$$r_{xy} = \frac{\sum_{k=1}^{m}(x_k-\bar{x})(y_k-\bar{y})}{\left[\sum_{k=1}^{m}(x_k-\bar{x})^2\right]^{1/2}\left[\sum_{k=1}^{m}(y_k-\bar{y})^2\right]^{1/2}}$$

Where $\bar{y} = \sum_{k=1}^{m}\frac{y_k}{m}$, $\bar{x} = \sum_{k=1}^{m}\frac{x_k}{m}$

จากนั้นจัดกลุ่มตามค่า coefficient ซึ่งมีค่าระหว่าง -1 ถึง 1 ค่าที่มีเข้าใกล้มากสุดจะมีผลต่อปริมาณผลผลิตสูงที่สุดด้วย

2) วิเคราะห์การพัฒนาปริมาณผลผลิต

เมื่อได้ตัวแปรที่สำคัญจากขั้นตอนก่อนหน้านี้แล้ว นำตัวแปรที่ได้มาแยกประเภทว่าเป็นปัญหาภายในโรงงานหรือนอกโรงงาน หากเป็นปัญหาภายในโรงงานให้พัฒนาปรับปรุง การขยับค่าเฉลี่ย (mean shift),และการลดค่าความแปรปรวน (variance reduction) แต่หากเป็นปัญหาภายนอกโรงงานให้พัฒนาการกระชับข้อจำกัด (Spec Tightening)ให้มากขึ้นโดยการสร้างเครื่องมือที่สามารถปรับค่าตัวแปรอย่างอัตโนมัติด้วยเทคนิค Genetic Algorithm (GA) ในการสร้างข้อจำกัดที่ดีที่สุด

จากงานวิจัยที่เกี่ยวข้องดังที่กล่าวมานั้น จะเห็นได้ว่าเทคนิคต้นไม้ตัดสินใจถูกนำมาใช้ในการหากฎเกณฑ์ที่จะเป็นแนวทางในการจำแนกชิ้นงานและทำการพัฒนาวิธีการผลิตเพื่อให้ได้ผลลัพธ์ที่ดีขึ้นจากแนวทางที่ได้มานี้ ซึ่งต้นไม้ตัดสินใจนั้นมีให้เลือกหลายอัลกอริทึม แต่อัลกอริทึม C4.5 เป็นอัลกอริทึมที่มีความสามารถในการจัดการข้อมูลที่เป็นค่าต่อเนื่องได้ดี และ มีการตัดเล็มต้นไม้ช่วยแก้ไขปัญหาข้อมูลเหมาะเกินไปซึ่งสามารถเกิดขึ้นได้กับข้อมูลขนาดใหญ่ เช่น โรงงานอุตสาหกรรม อีกทั้งยังนำค่าสหสัมพันธ์มาใช้ในการหาความสัมพันธ์กันระหว่างตัวแปรยังสามารถนำมาช่วยในการบอกแนวโน้มของค่าอื่นๆ ที่เกิดจากผลการเปลี่ยนแปลงของค่าตัวแปรใน

การทดลอง จึงทำให้สามารถทำการปรับตัวแปรได้โดยไม่ส่งผลกระทบต่อค่าอื่นๆ ดังนั้นงานวิจัยนี้จะใช้ต้นไม้ตัดสินใจอัลกอริทึม C4.5 มาทำการสร้างกฎเกณฑ์ที่จะใช้เป็นแนวโน้มในการปรับค่าพารามิเตอร์ที่เหมาะสมเพื่อสร้างแบบจำลองที่จะช่วยให้สามารถผลิตชิ้นงานที่มีคุณภาพดีได้มากขึ้น

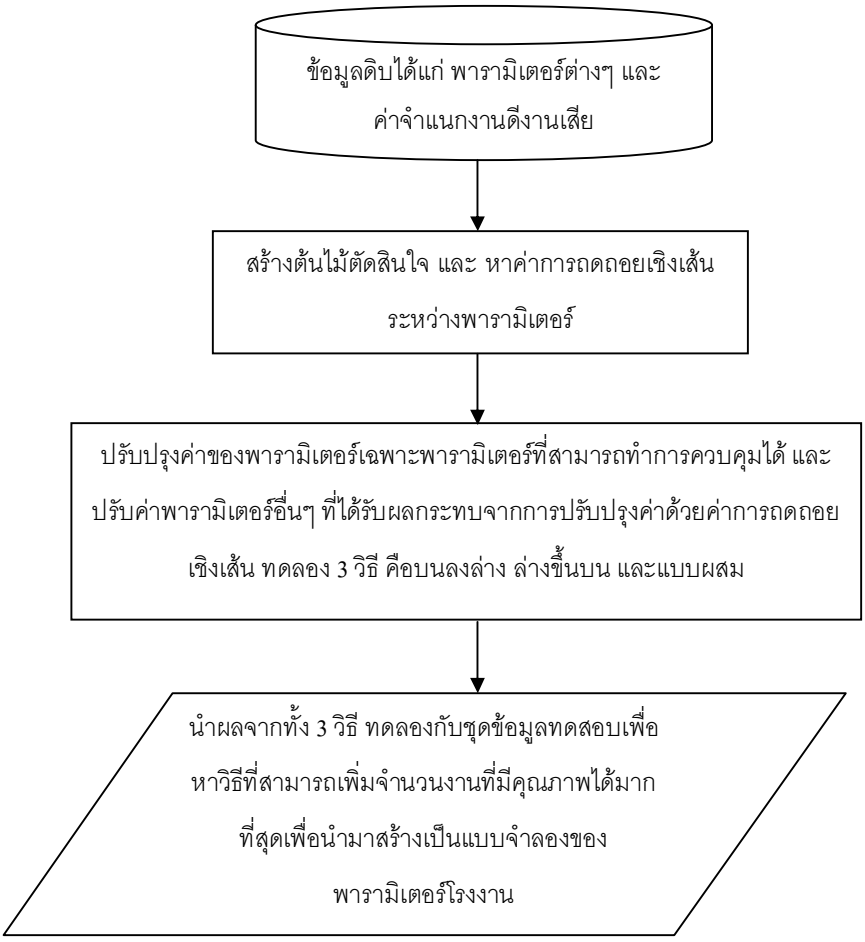
บทที่ 3

ระเบียบวิธีขั้นตอนที่เสนอ

ระเบียบขั้นตอนวิธี

งานวิจัยนี้เสนอแบบจำลองเพื่อใช้ในการพัฒนาคุณภาพของฮาร์ดดิสก์ให้มีการเพิ่มจำนวนงานที่มีคุณภาพมากขึ้น โดยจะนำเทคนิคต้นไม้ตัดสินใจมาใช้ในการเรียงลำดับความสำคัญของพารามิเตอร์และทำการทดลองปรับค่าพารามิเตอร์ของต้นไม้ตัดสินใจ 3 วิธี คือบนลงล่าง,ล่างขึ้นบน และแบบผสม เปรียบเทียบและหาวิธีที่ดีที่สุดที่จะใช้เป็นแบบจำลองในการปรับพารามิเตอร์ของโรงงาน

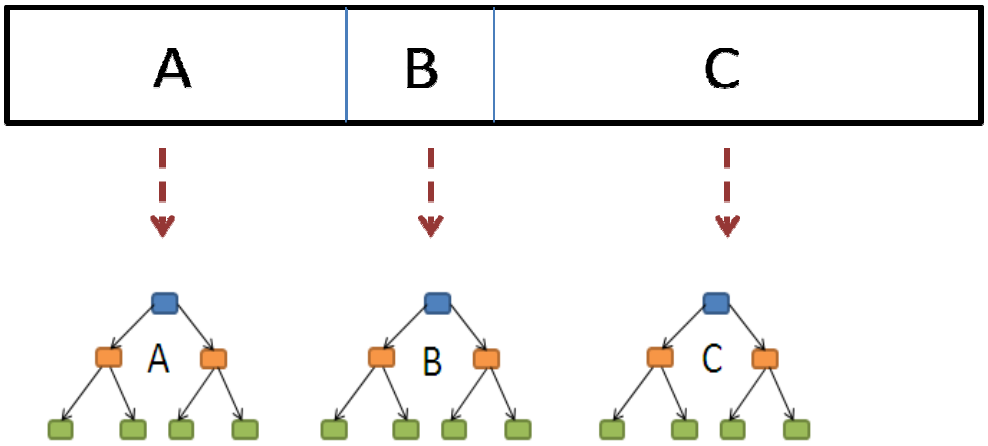
ขั้นตอนการทำงานของวิธีที่นำเสนอภาพรวมมีดังแสดงรายละเอียดใน*ภาพที่ 3-1*



ภาพที่ 3-1 ขั้นตอนการพัฒนาคุณภาพการผลิตฮาร์ดดิสก์

ขั้นตอนวิธีการทดลองมีการแบ่งข้อมูลออกเป็นส่วนๆ ดังแสดงรายละเอียดดังนี้

1. ข้อมูลดิบของโรงงานถูกแบ่งออกเป็น 3 ส่วน คิดเป็นสัดส่วนของ A : B : C เป็น 4 : 1 : 5 ข้อมูลในแต่ละส่วนถูกนำมาสร้างต้นไม้ตัดสินใจ



ภาพที่ 3-2 วิธีการแบ่งข้อมูลในการทดลอง

โดยที่ต้นไม้ตัดสินใจ A ใช้ในการทดลองปรับพารามิเตอร์ทั้ง 3 แบบ คือ Top-down,ล่างขึ้นบน และแบบผสม

ต้นไม้ตัดสินใจ B ใช้ในการทดสอบการปรับพารามิเตอร์ของต้นไม้ตัดสินใจ A ว่าควรหยุดทำการปรับเมื่อไหร่ กล่าวคือ ให้หยุดทำการปรับเมื่อใช้ข้อมูลที่ทำการปรับพารามิเตอร์กับกฎเกณฑ์ของต้นไม้ตัดสินใจ B แล้วไม่สามารถทำให้จำนวนงานดีเพิ่มขึ้น

ต้นไม้ตัดสินใจ C ใช้เป็นโรงงานจำลองเพื่อตรวจสอบวิธีการปรับพารามิเตอร์ว่าสามารถใช้ได้ดีกับกฎเกณฑ์อื่นๆ ทั่วไปหรือไม่

2. ทำการทดลองปรับต้นไม้ตัดสินใจ A และนำวิธีการปรับที่ได้ไปแก้ไขข้อมูลดิบ B และทดสอบกฎเกณฑ์ของต้นไม้ตัดสินใจ B เพื่อทำการวัดผลประสิทธิภาพของการปรับพารามิเตอร์ และใช้เป็นตัวหยุดการปรับพารามิเตอร์
3. เมื่อทำการทดลองจนสิ้นสุดการปรับพารามิเตอร์และได้วิธีการปรับพารามิเตอร์แล้ว นำวิธีการปรับที่ได้ไปแก้ไขข้อมูลดิบ C และทดสอบกฎเกณฑ์ของต้นไม้ตัดสินใจ C เพื่อทดสอบว่าสามารถทำการปรับปรุงได้เท่าไหร่ โดยการทดสอบนี้เปรียบเสมือนการทดสอบกับโรงงานจำลอง

4. ทำการปรับพารามิเตอร์ของต้นไม้ตัดสินใจ A ตามวิธีข้อที่ 3 โดยทดลองวิธีการปรับพารามิเตอร์ บนลงล่าง (Top-down) ล่างขึ้นบน (Bottom-up) และแบบผสม (Hybrid) จากนั้นนำผลที่ได้มาทำการเปรียบเทียบเพื่อหาวิธีที่ดีที่สุดที่จะเพิ่มจำนวนงานดี

3.1 การทำความสะอาดข้อมูล

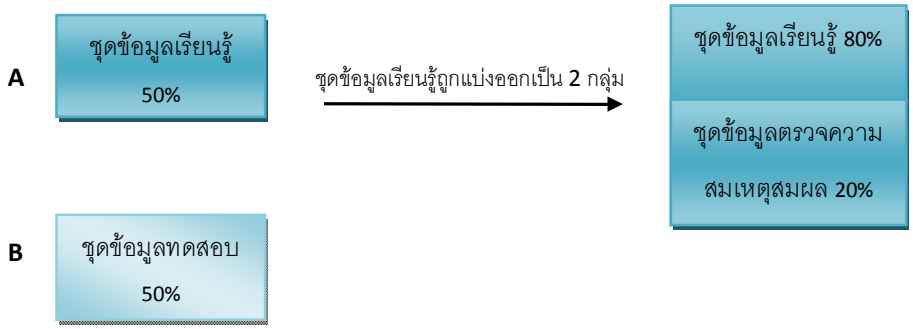
ข้อมูลที่นำมาใช้ในการทดลองเป็นข้อมูลการผลิตฮาร์ดดิสก์ชนิดหนึ่งใน 1 วัน ขั้นตอนนี้จึงทำการตรวจสอบข้อมูลทั้งหมดก่อนการทดลอง ดังต่อไปนี้

3.1.1 ค่าพารามิเตอร์ต้องไม่ใช่ค่าว่าง เพราะจะทำให้เกิดความผิดพลาดในการสร้างต้นไม้ตัดสินใจได้ จึงต้องทำการลบข้อมูลทั้งแถวที่มีค่าว่างนั้นออกไป

3.1.2 ค่าพารามิเตอร์นั้นต้องไม่ใช่ค่าผิดพลาด จากการวัด หรือ การบันทึกผลจากการวัด เพราะจะทำให้ต้นไม้ตัดสินใจไม่ถูกต้องได้ จึงจำเป็นต้องทำการลบข้อมูลทั้งแถวที่มีค่าผิดพลาดนั้นออกไป

3.2 การแบ่งข้อมูล

เนื่องจากการทดลองนี้ไม่สามารถทำการปรับพารามิเตอร์ในโรงงานได้จริง เพราะอาจก่อให้เกิดความเสียหายในการผลิตได้ จึงทำการสร้างโรงงานจำลอง 2 โรงงานด้วยต้นไม้ตัดสินใจทำการปรับปรุงโรงงานที่ 1 นำผลการปรับปรุงที่ได้มาใช้กับโรงงานที่ 2 เพื่อทดสอบวิธีการปรับปรุงว่าให้ผลได้ดีกับทุกข้อมูลหรือไม่ งานวิจัยนี้จะทำการแบ่งข้อมูลออกเป็น 3 ส่วน ดังแสดงรายละเอียดใน*ภาพที่ 3-3*



ภาพที่ 3-3 สัดส่วนการแบ่งข้อมูลการผลิต

จาก*ภาพที่ 3-3* ข้อมูลการผลิตทั้งหมดที่นำมาทดลองคือ ข้อมูล A และข้อมูล B ซึ่งถูกแบ่งออกเป็น 2 กลุ่ม เท่าๆ กัน กลุ่ม A เป็นชุดข้อมูลเรียนรู้ หรือจำลองเป็นโรงงานที่ 1 และกลุ่ม B เป็นชุดข้อมูลทดสอบ หรือจำลองเป็นโรงงานที่ 2 ในกลุ่ม A ถูกแบ่งย่อยๆ อีกเป็น 2 กลุ่ม กลุ่มแรกร้อยละ 80 เป็นชุดข้อมูลเรียนรู้ ใช้ในการปรับพารามิเตอร์ในแต่ละโหนดของต้นไม้ตัดสินใจ และกลุ่มร้อยละ 20 เป็นชุดข้อมูลตรวจความสมเหตุสมผล เพื่อตรวจสอบชุดข้อมูลเรียนรู้ร้อยละ 80 ในการหยุดทำการปรับพารามิเตอร์

3.3 **สร้างต้นไม้ตัดสินใจด้วยอัลกอริทึม** C4.5

นำข้อมูลทั้ง 3 กลุ่มจาก 3.2 มาสร้างต้นไม้ตัดสินใจอัลกอริทึม C4.5 โดยโปรแกรม Weka เวอร์ชัน 3.6.0 โดยเลือก Classify ด้วย Tree J48 เพื่อให้ได้กฎของการจำแนกกลุ่มคุณภาพของผลิตภัณฑ์จากการเรียนรู้จากข้อมูลทั้ง 3 กลุ่ม

3.4 **ปรับพารามิเตอร์แบบ บนลงล่าง** (Top-down) **ล่างขึ้นบน** (Bottom-up) **และแบบผสม** (Hybrid)

เพื่อความเข้าใจในอัลกอริทึมจึงมีการกำหนดนิยามของข้อมูลที่ใช้ในวิธีที่เสนอมีรายละเอียดดังต่อไปนี้

นิยาม 3.1 T คือ ต้นไม้ตัดสินใจที่ได้จากกลุ่มข้อมูล A ร้อยละ 80 ที่เป็นชุดข้อมูลเรียนรู้

นิยาม 3.2 V คือ กลุ่มข้อมูล A ร้อยละ 20 ที่เป็นชุดข้อมูลตรวจความสมเหตุสมผล

นิยาม 3.3 T' คือ ต้นไม้ตัดสินใจที่ได้จากกลุ่มข้อมูล A ร้อยละ 20 ที่เป็นชุดข้อมูลตรวจความสมเหตุสมผล

การปรับพารามิเตอร์จะทำการปรับเฉพาะพารามิเตอร์ที่สามารถควบคุมได้เท่านั้น โดยการปรับต้องทำการหาค่าสหสัมพันธ์กับพารามิเตอร์ทั้งหมดก่อนเพื่อสามารถทำการปรับพารามิเตอร์ได้โดยไม่ส่งกระทบพารามิเตอร์อื่นๆหากค่าพารามิเตอร์ที่ทำการปรับนั้นมีค่าสหสัมพันธ์กับพารามิเตอร์อื่นๆ ให้นำค่าการถดถอยเชิงเส้นในการหาค่าพารามิเตอร์นั้นๆ

3.4.1 การปรับพารามิเตอร์แบบบนลงล่าง

การปรับพารามิเตอร์แบบบนลงล่าง เริ่มต้นทำการปรับที่โหนดรากของต้นไม้ตัดสินใจมีอัลกอริทึม ดังแสดงรายละเอียดใน*ภาพที่ 3-4*

```
TOP-DOWN[T,V,T']

Input: T, a decision tree
       V, a validation set
       T', a decision tree constructed from the
          validation set

Output: L, a list of pairs of parameter and its value

Initialize: i ← 0
1  WHILE(Accuracy is increasing)
2      N_i ← extract all nodes in level i^th of T
3      FOR EACH N_i
4          C_j ← all children of N_i
4          IF N_i is controllable parameter
5              IF C_j is PASS or FAIL
6                  Add a pair of parameter and its value in the parent of C_j to L.
7      Accuracy ← EVALUATE[V,T',L]
8      i ← i+1
9  RETURN L
```

```
EVALUATE[S,T,L]

Input: S, sample set
       T, a decision tree
       L, a list of adjustable parameter

Output: Accuracy of S evaluated on T
```

1 *S'* ←Adjust the value in *S* using all dependent parameters from *L* and linear regression
2 *Accuracy* ← Evaluate the accuracy of *S* ' on *T*
3 RETURN *Accuracy*

ภาพที่ 3-4 อัลกอริทึมบนลงล่าง

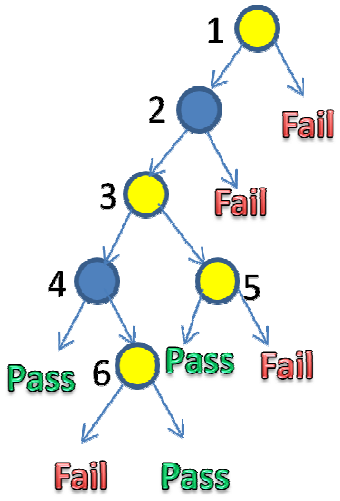
การปรับพารามิเตอร์แบบบนลงล่าง จะทำการปรับพารามิเตอร์จากบนลงล่าง กล่าวคือทำการปรับจาก ราก ไปสู่ โหนดระดับถัดไป

ตัวอย่างที่ 3.1 การปรับแบบบนลงล่าง

กำหนดให้ วงกลมสีเหลืองคือพารามิเตอร์ที่สามารถทำการปรับค่าได้

วงกลมสีฟ้าคือพารามิเตอร์ที่ไม่สามารถทำการปรับค่าได้

ตัวอย่างต้นไม้ตัดสินใจ1



ภาพที่ 3-5 ตัวอย่างต้นไม้ตัดสินใจ1

เริ่มต้นที่โหนดรากก่อน คือ เริ่มที่โหนดหมายเลข 1 ซึ่งเป็นพารามิเตอร์ที่สามารถปรับค่าได้ บันทึกค่าที่จะทำให้งานไม่ตกไปกิ่ง Fail ทดลองทำการปรับค่านี้ซึ่งการปรับค่านั้นต้องคำนึงถึงค่าสหสัมพันธ์กับพารามิเตอร์อื่นด้วยโดยใช้สมการเชิงเส้นถดถอยในการปรับพารามิเตอร์ที่ผลกระทบในกลุ่มร้อยละ 20 ที่เป็นชุดข้อมูลตรวจความสมเหตุสมผล หากพบว่า ไม่ได้จำนวน

งานดีเพิ่มขึ้นให้หยุด แต่ถ้าได้จำนวนงานดีเพิ่มขึ้นให้ดำเนินการต่อ นั่นคือเลื่อนมาที่โหนดหมายเลข 2 แต่โหนดหมายเลข 2 เป็นพารามิเตอร์ที่ไม่สามารถปรับค่าได้ จึงข้ามทำโหนดถัดไปนั่นคือโหนดหมายเลข 3 ซึ่งเป็นพารามิเตอร์ที่สามารถปรับค่าได้ แต่โหนดหมายเลข 3 ไม่สามารถระบุว่าเป็นงานดีหรืองานเสียจึงข้ามไปทำโหนดถัดไป ซึ่งมี 2 ทาง คือ โหนดหมายเลข 4 และ โหนดหมายเลข 5 ในโหนดหมายเลข 4 เป็นพารามิเตอร์ที่ไม่สามารถปรับค่าได้จึงต้องเลื่อนไปโหนดหมายเลข 6 ซึ่งเป็นพารามิเตอร์ที่สามารถปรับค่าได้ ทำการปรับค่าด้วยวิธีการเดียวกับการปรับโหนดหมายเลข 1 หากให้จำนวนงานดีเพิ่มขึ้นทำการระบุค่า ถัดมาในโหนดหมายเลข 5 เป็นพารามิเตอร์ที่สามารถปรับค่าได้ ทำการปรับค่าด้วยวิธีการเช่นเดียวกัน หากให้จำนวนงานดีเพิ่มขึ้นทำการระบุค่าเช่นกัน

หากตัวอย่างนี้ สามารถทำให้จำนวนงานดีเพิ่มขึ้นได้หมด ผลลัพธ์ที่ได้ คือ ผลลัพธ์จากการปรับโหนดหมายเลข 1 ผสานกับ ผลลัพธ์จากการปรับโหนดหมายเลข 5 ผสานกับ ผลลัพธ์จากการปรับโหนดหมายเลข 6

3.4.2 การปรับพารามิเตอร์แบบล่างขึ้นบน

การปรับพารามิเตอร์แบบล่างขึ้นบน เริ่มต้นทำการปรับที่โหนดใบของต้นไม้ตัดสินใจ มีอัลกอริทึม ดังแสดงรายละเอียดใน*ภาพที่ 3-6*

BOTTOM-UP[*T,V,T'*]

Input: *T,* a decision tree
V, a validation set
T', a decision tree constructed from the validation set

Output: *L,* a list of pairs of parameter and its value

Initialize: i ← depth of *T*
1 WHILE(*Accuracy* is increasing)
2 N_i ← extract all nodes in level i^{th} of *T*
3 FOR EACH N_i
4 P_j ← all parent nodes of N_i
4 IF P_i is controllable parameter
5 IF P_j is PASS or FAIL
6 Add a pair of parameter and its value in P_j to *L*.
7 *Accuracy* ← EVALUATE[*V,T',L*]
8 i ← i – 1
9 RETURN *L*

EVALUATE[*S,T,L*]

Input: *S*, sample set
T, a decision tree
L, a list of adjustable parameter

Output: Accuracy of *S* evaluated on *T*

1 *S'* ←Adjust the value in *S* using all dependent parameters from *L* and linear regression

2 *Accuracy* ← Evaluate the accuracy of *S* ‘ on *T*
3 RETURN *Accuracy*

ภาพที่ 3-6 อัลกอริทึมล่างขึ้นบน

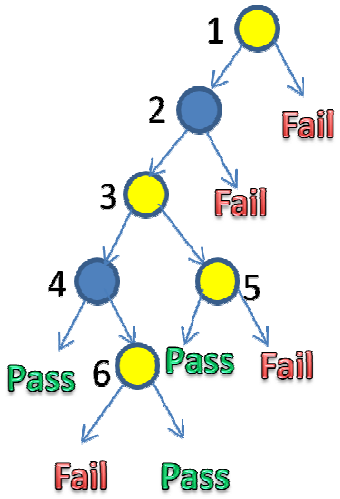
การปรับพารามิเตอร์แบบล่างขึ้นบน จะทำการปรับพารามิเตอร์จากบนลงล่าง กล่าวคือทำการปรับจาก โหนดที่ให้ลูกหลานเป็นใบ ไป โหนดถัดไปที่ระดับน้อยกว่า

ตัวอย่างที่ 3.2 การปรับแบบล่างขึ้นบน

กำหนดให้ วงกลมสีเหลืองคือพารามิเตอร์ที่สามารถทำการปรับค่าได้

วงกลมสีฟ้าคือพารามิเตอร์ที่ไม่สามารถทำการปรับค่าได้

ตัวอย่างต้นไม้ตัดสินใจ2



ภาพที่ 3-7 ตัวอย่างต้นไม้ตัดสินใจ2

เริ่มต้นที่โหนดสุดท้ายระดับต่ำสุดก่อน นั่นคือโหนดหมายเลข 5 และ โหนดหมายเลข 6 เริ่มที่โหนดหมายเลข 6 ซึ่งเป็นพารามิเตอร์ที่สามารถปรับค่าได้ บันทึกค่าที่จะทำให้งานไม่ตกไปกิ่ง Fail ทดลองทำการปรับค่านี้ซึ่งการปรับค่านั้นต้องคำนึงถึงค่าสหสัมพันธ์กับพารามิเตอร์อื่นด้วยโดยใช้สมการเชิงเส้นถดถอยในการปรับพารามิเตอร์ที่ผลกระทบในกลุ่มร้อยละ 20ที่เป็นชุดข้อมูลตรวจความสมเหตุสมผล หากพบว่า ไม่ได้จำนวนงานดีเพิ่มขึ้นให้หยุด แต่ถ้าได้จำนวนงานดีเพิ่มขึ้นให้ดำเนินการต่อนั่นคือเลื่อนขึ้นมาที่โหนดหมายเลข 4 แต่โหนดหมายเลข 4

เป็นพารามิเตอร์ที่ไม่สามารถปรับค่าได้ จึงเลื่อนระดับขึ้นไปอีกที่โหนดหมายเลข 3 แต่โหนดหมายเลข 3 ไม่สามารถระบุว่าเป็นงานดีหรืองานเสียจึงข้ามไปทำโหนดถัดไป คือโหนดหมายเลข 2 แต่โหนดหมายเลข 2 เป็นพารามิเตอร์ที่ไม่สามารถปรับค่าได้ จึงเลื่อนระดับขึ้นไปอีก เป็นโหนดหมายเลข 1 ซึ่งเป็นพารามิเตอร์ที่สามารถปรับค่าได้ ทำการปรับด้วยวิธีการเดียวกัน เป็นการจบกระบวนการของการปรับที่เริ่มจากโหนดหมายเลข 6 ต่อมาที่โหนดหมายเลข 5 เป็นพารามิเตอร์ที่สามารถปรับค่าได้ ทำการปรับเช่นเดียวกัน หากได้จำนวนงานดีเพิ่มขึ้นให้เลื่อนโหนดระดับขึ้นไป คือ โหนดหมายเลข แต่โหนดหมายเลข 3 ไม่สามารถระบุว่าเป็นงานดีหรืองานเสียจึงข้ามไปทำโหนดหมายเลข 2 แต่โหนดหมายเลข 2 เป็นพารามิเตอร์ที่ไม่สามารถปรับค่าได้ จึงเลื่อนไปเป็นโหนดหมายเลข 1ทำการปรับค่า

หากตัวอย่างนี้ สามารถทำให้จำนวนงานดีเพิ่มขึ้นได้หมด ผลลัพธ์ที่ได้ คือ ผลลัพธ์ที่ได้จากการปรับที่เริ่มจากโหนดหมายเลข 6 ผสานกับ ผลลัพธ์ที่ได้จากการปรับที่เริ่มจากโหนดหมายเลข 5

3.4.3 การปรับพารามิเตอร์แบบผสม

การปรับพารามิเตอร์แบบผสม เป็นการรวมผลที่ได้จากการปรับพารามิเตอร์แบบบนลงล่าง และล่างขึ้นบน มีอัลกอริทึม ดังแสดงรายละเอียดใน*ภาพที่ 3-8*

HYBRID[*T,V,T'*]

Input: *T,* a decision tree
V, a validation set
T', a decision tree constructed from the validation set

Output: *L,* a list of pairs of parameter and its value

1 $L_1 \leftarrow$ TOP-DOWN[*T,V,T'*]

2 $L_2 \leftarrow$ BOTTOM-UP[*T,V,T'*]

3 RETURN $L_1 \cup L_2$

ภาพที่ 3-8 **อัลกอริทึมแบบผสม**

กล่าวคือนำผลลัพธ์ที่ได้จากวิธีการแบบบนลงล่าง และล่างขึ้นบน มาทำการผสานกันเพื่อให้ได้วิธีการปรับพารามิเตอร์ทั้งหมดที่ครอบคลุมที่สุด

จากนั้นนำผลการทดลองที่ได้จากการปรับพารามิเตอร์ทั้ง 3 วิธี ไปทดสอบกับชุดข้อมูลทดสอบ หรือ ชุดข้อมูลกลุ่ม B เพื่อสรุปวิธีที่สามารถเพิ่มจำนวนงานดีได้มากที่สุดและนำไปใช้ในการสร้างแบบจำลองในการพัฒนาคุณภาพของกระบวนการผลิตฮาร์ดดิสก์ต่อไป

บทที่ 4

การทดลองและผลการทดลอง

จากวิธีที่นำเสนอนั้นจะทำการทดลองปรับค่าพารามิเตอร์ 3 แบบ คือบนลงล่าง ล่างขึ้นบน และ แบบผสม ด้วยอัลกอริทึมที่นำเสนอมา เพื่อนำผลที่ได้มาเปรียบเทียบหาวิธีการปรับพารามิเตอร์ที่ดีที่สุด

ในการทดลองจะแบ่งออกเป็นสองส่วนได้แก่ ส่วนการสร้างต้นไม้ตัดสินใจเพื่อหาพารามิเตอร์ที่สำคัญ และส่วนของการปรับพารามิเตอร์ที่สามารถควบคุมได้ โดยมีรายละเอียดต่างๆ รวมไปถึงเครื่องมือที่ใช้ในการทดลองและการวิเคราะห์ผลการทดลองดังต่อไปนี้

4.1 เครื่องมือที่ใช้

การทดลองนี้ได้ทดลองโดยใช้เครื่อง Intel T2130 1.86 GHz หน่วยความจำ 2048 MB และใช้ โปรแกรม Weka version 3.6.0 เป็นเครื่องมือในการสร้างต้นไม้ตัดสินใจ และ MS Excel 2007 เป็นเครื่องมือในการคำนวณข้อมูลการปรับค่าพารามิเตอร์

4.2 ข้อมูลที่นำมาใช้ในการทดลอง

ข้อมูลที่นำมาใช้เป็นข้อมูลการผลิตฮาร์ดดิสก์ 3 ชนิดใน 1 วัน และในแต่ละชนิดมีพารามิเตอร์การผลิตที่แตกต่างกัน แบ่งออกเป็น 3 ประเภท คือ พารามิเตอร์ที่สามารถควบคุมได้ , พารามิเตอร์ที่มีความสัมพันธ์กับพารามิเตอร์ที่สามารถควบคุมได้ และพารามิเตอร์ที่ไม่สามารถควบคุมได้ ดังมีภาพและรายละเอียดแสดงใน*ตารางที่ 4-1*

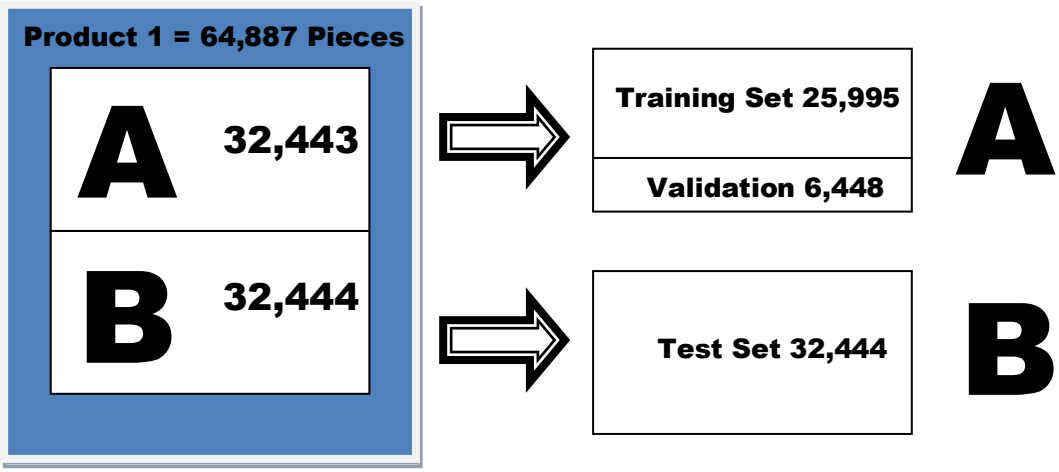
ผลิตภัณฑ์	พารามิเตอร์ที่สามารถควบคุมได้	พารามิเตอร์ที่มีความสัมพันธ์กับพารามิเตอร์ที่สามารถควบคุมได้	พารามิเตอร์ที่ไม่สามารถควบคุมได้
Product 1	P1,P2,P3,P4,P5,P6	P7, P8	P9,P10,P11
Product 2	P2,P3,P6	P4,P5,P8,P9	P1,P7
Product 3	P1,P2,P3,P8	P4,P5,P6,P7,P9	-

ตารางที่ 4-1 ข้อมูลการผลิตฮาร์ดดิสก์ที่นำมาใช้ในการทดลอง

4.3 การทดลองปรับพารามิเตอร์ของ Product 1

4.3.1 การแบ่งข้อมูล

งานวิจัยนี้ต้องการแบ่งข้อมูลการผลิตฮาร์ดดิสก์ในหนึ่งวันออกเป็นสองส่วนเท่ากัน ส่วนที่หนึ่งแบ่งออกเป็นชุดข้อมูลเรียนรู้ร้อยละ 80 และชุดข้อมูลตรวจความสมเหตุสมผลอีกร้อยละ 20 และในส่วนที่สองใช้เป็นชุดข้อมูลทดสอบ สัดส่วนการแบ่งข้อมูลแสดงราย ละเอียดใน*ภาพที่ 4-1*



ภาพที่ 4-1 สัดส่วนของในการแบ่งข้อมูลของ Product 1

4.3.2 การสร้างต้นไม้ตัดสินใจ

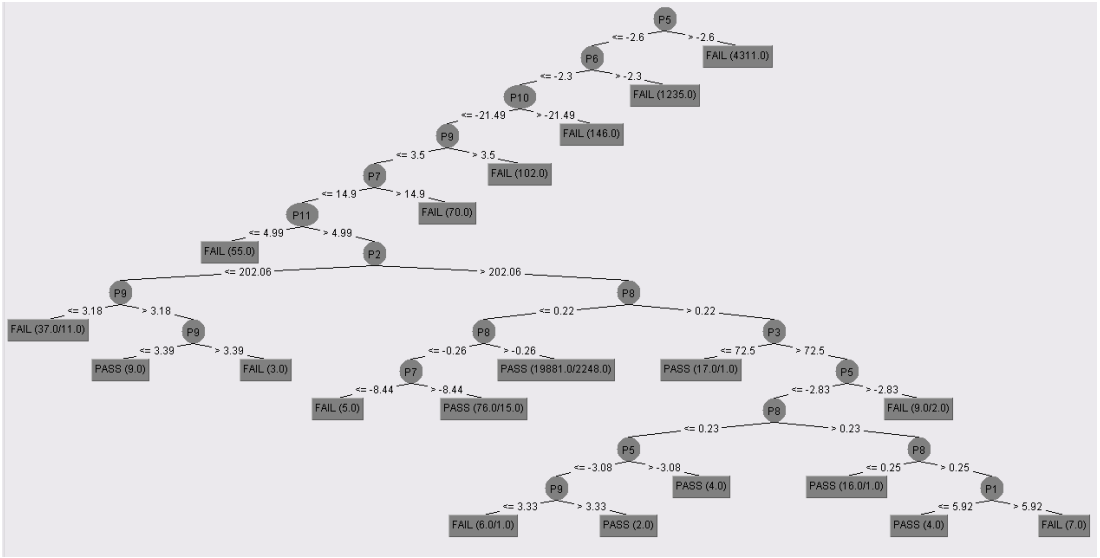
ในการทดลองนี้ใช้ต้นไม้ตัดสินใจในการหาพารามิเตอร์ที่สำคัญที่มีผลต่อคุณภาพการผลิตฮาร์ดดิสก์ อัลกอริทึมที่นำมาใช้คือ C4.5 ซึ่งทำการสร้างต้นไม้ตัดสินใจด้วยโปรแกรม Weka เลือก Classify ด้วย J48 –C 0.25 M 2 ทำการสร้างต้นไม้ตัดสินใจ 3 ต้นจากชุดข้อมูลเรียนรู้ร้อยละ 80 และชุดข้อมูลตรวจความสมเหตุสมผลอีกร้อยละ 20 ในส่วนแบ่งแรกและชุดข้อมูลทดสอบในส่วนที่สอง

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลเรียนรู้ร้อยละ 80 จำนวน 25,995 ตัว ด้วย Weka ใช้เวลาการคำนวณ 9.13 วินาทีและให้ความถูกต้องร้อยละ 91.04 ซึ่งแสดงรายละเอียดใน *ภาพที่ 4-2* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-3*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       23666               91.0406 %
Incorrectly Classified Instances      2329                8.9594 %
Kappa statistic                          0.7766
Mean absolute error                      0.1569
Root mean squared error                  0.2818
Relative absolute error                 36.2379 %
Root relative squared error             60.5596 %
Total Number of Instances            25995
```

ภาพที่ 4-2 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 1



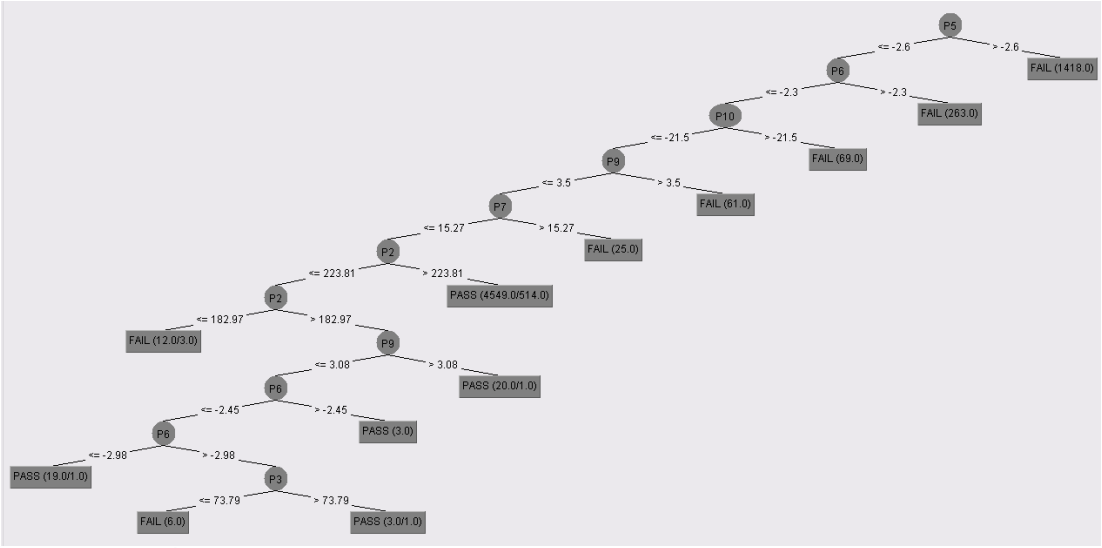
ภาพที่ 4-3 **ต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 1

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 จำนวน 6,488 ตัว ด้วย Weka ใช้เวลาการคำนวณ 0.63 วินาทีและให้ความถูกต้องร้อยละ 91.56 ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-4* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-5*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        5904               91.5633 %
Incorrectly Classified Instances       544                8.4367 %
Kappa statistic                          0.8099
Mean absolute error                      0.1449
Root mean squared error                  0.2726
Relative absolute error                 31.1688 %
Root relative squared error             56.5413 %
Total Number of Instances             6448
```

ภาพที่ 4-4 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 1



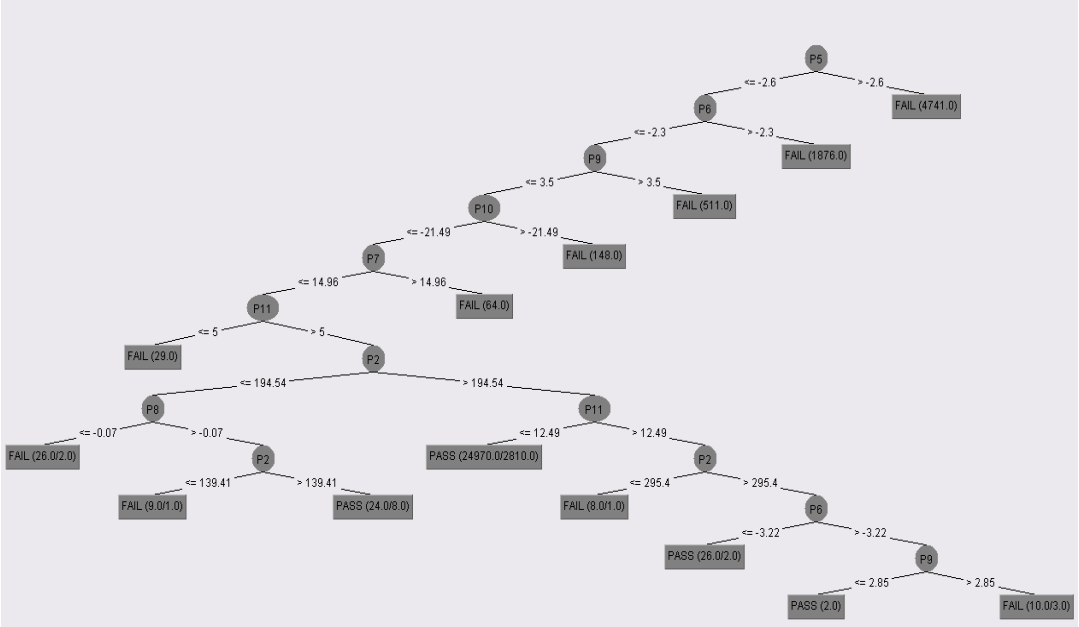
ภาพที่ 4-5 **ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 1

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลทดสอบ จำนวน 32,444 ตัว ด้วย Weka ใช้เวลาการคำนวณ 11.72 วินาทีและให้ความถูกต้องร้อยละ 91.18 ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-6* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-7*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       29583               91.1817 %
Incorrectly Classified Instances      2861                8.8183 %
Kappa statistic                          0.7795
Mean absolute error                      0.1553
Root mean squared error                  0.2796
Relative absolute error                 35.9548 %
Root relative squared error             60.1724 %
Total Number of Instances            32444
```

ภาพที่ 4-6 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 1



ภาพที่ 4-7 **ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 1

4.3.3 การปรับพารามิเตอร์

ปรับพารามิเตอร์ทำที่ต้นไม้ตัดสินใจของชุดข้อมูลเรียนรู้ร้อยละ 80 ปรับเฉพาะพารามิเตอร์ที่สามารถทำการปรับได้เท่านั้น ได้แก่ P1-P8 ทำการทดลองปรับพารามิเตอร์ 3 แบบ คือบนลงล่าง,ล่างขึ้นบน และ แบบผสม โดยมีเงื่อนไขว่า เมื่อทำการปรับแล้วไม่สามารถทำ ให้ชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 มีจำนวนงานดีมากขึ้นจะหยุดทำการปรับทันที

4.3.3.1 การปรับพารามิเตอร์แบบบนลงล่าง

การปรับพารามิเตอร์เริ่มต้นที่โหนดรากของต้นไม้ตัดสินใจ ผลจากการปรับแสดงรายละเอียดใน*ตารางที่ 4-2*

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผล ร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
P5 <= -2.6	4,777
P6 <= -2.3	5,097
P7<= 14.9 มีผลทำให้ P2 = 178.32 และ P4 = 8070.41	0 (หยุดทำการปรับ)

ตารางที่ 4-2 การปรับข้อมูลของ Product 1 แบบบนลงล่าง

การปรับพารามิเตอร์แบบบนลงล่าง ของ Product 1 สามารถทำได้โดยการปรับ P5<=-2.6 และ P6 <= -2.3

4.3.3.2 การปรับพารามิเตอร์แบบล่างขึ้นบน

เริ่มต้นที่โหนดสุดท้ายของต้นไม้ตัดสินใจ ซึ่งมีด้วยกัน 4 โหนด ผลจากการปรับแต่ละโหนดแสดงรายละเอียดใน*ตารางที่ 4-3, ตารางที่ 4-4, ตารางที่ 4-5 และ ตารางที่ 4-6* ตามลำดับ

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผล ร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
P7 <= 14.9 มีผลทำให้ P2 = 178.32 และ P4 = 8070.41	0 (หยุดทำการปรับ)

ตารางที่ 4-3 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 1

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
P7 > -8.44 มีผลให้ P2 = 487.92 และ P4 = 12901.97	4,637
P8 > -0.26 มีผลให้ P3 = 73.82	4,637 (หยุดทำการปรับ)

ตารางที่ 4-4 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 2

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
P5 > -3.08	4,777
P5 <= -2.83	4,777 (หยุดทำการปรับ)

ตารางที่ 4-5 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 3

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
P1 <= 5.92	4,594

ตารางที่ 4-6 การปรับข้อมูลของ Product 1 แบบล่างขึ้นบน ที่โหนดที่ 4

การปรับพารามิเตอร์แบบล่างขึ้นบน ของ Product 1 สามารถทำได้โดยการปรับ P7 > -8.44 และ P5 > -3.08

4.3.3.3 การปรับพารามิเตอร์แบบผสม

ทำการรวมผลที่ได้จากการปรับพารามิเตอร์แบบบนลงล่าง และล่างขึ้นบน จะได้ -3.08< P5 <=-2.6 และ P6 <= -2.3 และ P7 > -8.44

4.3.3.4 เปรียบเทียบบนลงล่าง,ล่างขึ้นบน และแบบผสม

ทำการเปรียบเทียบทั้ง 3 วิธีด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20 ซึ่งแสดงรายละเอียดใน*ตารางที่ 4-7* และ ชุดข้อมูลทดสอบซึ่งแสดงรายละเอียดใน*ตารางที่ 4-8*

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,594
ปรับแบบบนลงล่าง	5,097
ปรับแบบล่างขึ้นบน	4,833
ปรับแบบผสม	5,194

ตารางที่ 4-7 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 1 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20

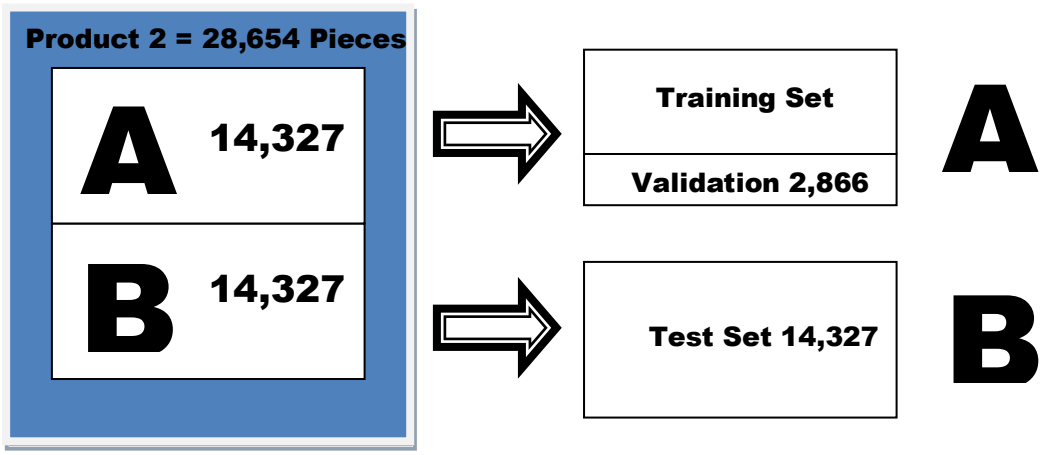
กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลทดสอบ**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	25,032
ปรับแบบบนลงล่าง	26,912
ปรับแบบล่างขึ้นบน	25,662
ปรับแบบผสม	27,266

ตารางที่ 4-8 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 1 ด้วยชุดข้อมูลทดสอบ

4.4 การทดลองปรับพารามิเตอร์ของ Product 2

4.4.1 การแบ่งข้อมูล

แบ่งข้อมูลออกเป็น 3 ส่วน เช่นเดิมกับการทดลอง Product 1 สัดส่วนการแบ่งข้อมูลแสดงรายละเอียดใน*ภาพที่ 4-8*



ภาพที่ 4-8 **สัดส่วน ในการแบ่งข้อมูลของ** Product 2

4.4.2 การสร้างต้นไม้ตัดสินใจ

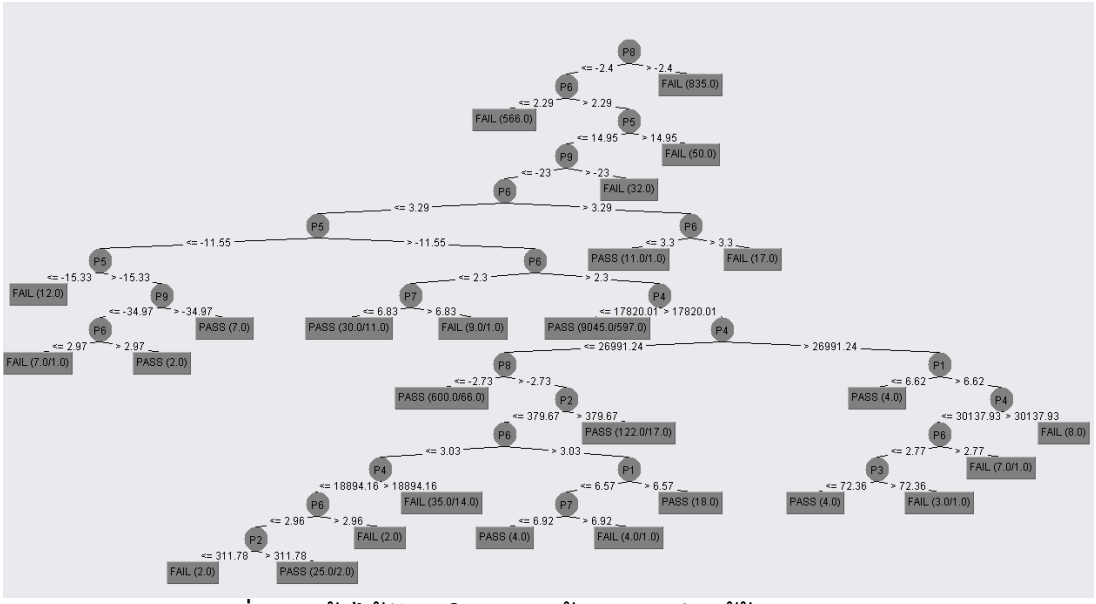
ในการทดลองนี้ใช้ต้นไม้ตัดสินใจในการหาพารามิเตอร์ที่สำคัญที่มีผลต่อคุณภาพการผลิตฮาร์ดดิสก์ อัลกอริทึมที่นำมาใช้คือ C4.5 ซึ่งทำการสร้างต้นไม้ตัดสินใจด้วยโปรแกรม Weka เลือก Classify ด้วย J48 –C 0.25 M 2 ทำการสร้างต้นไม้ตัดสินใจ 3 ต้นจากชุดข้อมูลเรียนรู้ร้อยละ 80 และชุดข้อมูลตรวจความสมเหตุสมผลอีกร้อยละ 20 ในส่วนแบ่งแรกและชุดข้อมูลทดสอบในส่วนที่สอง

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลเรียนรู้ร้อยละ 80 จำนวน 11,461 ตัว ด้วย Weka ใช้เวลาการคำนวณ 1.92 วินาทีและให้ความถูกต้องร้อยละ 93.29 ซึ่งแสดงรายละเอียดใน *ภาพที่ 4-9* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4*-10

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       10693               93.299  %
Incorrectly Classified Instances       768                6.701  %
Kappa statistic                          0.7608
Mean absolute error                      0.1188
Root mean squared error                  0.2482
Relative absolute error                 37.4613 %
Root relative squared error             62.3517 %
Total Number of Instances            11461
```

ภาพที่ 4-9 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 2



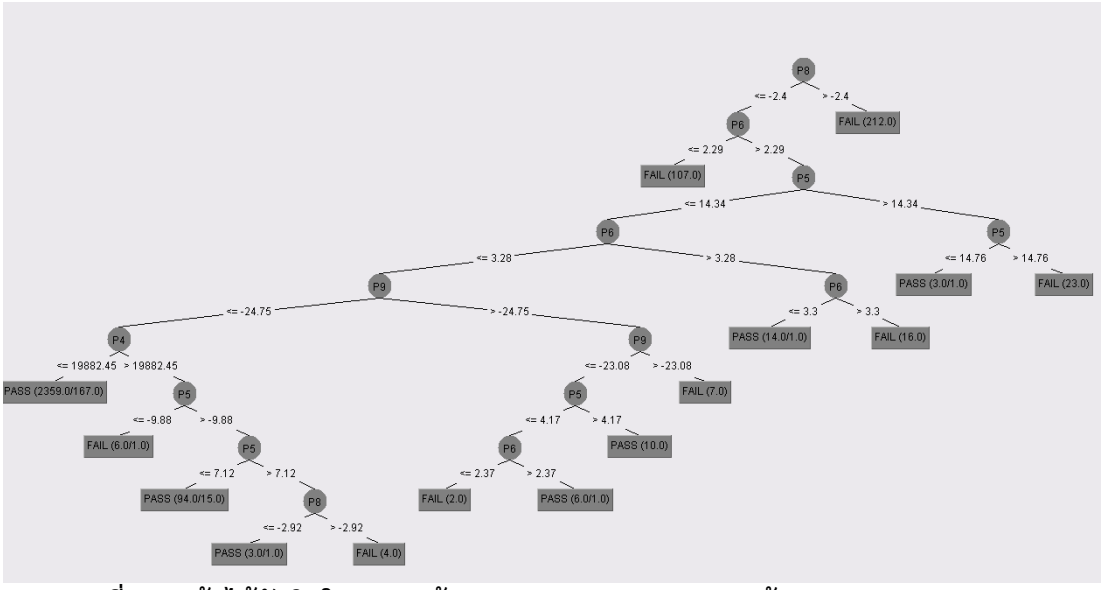
ภาพที่ 4-10 **ต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 2

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 จำนวน 2,866 ตัว ด้วย Weka ใช้เวลาการคำนวณ 0.19 วินาทีและให้ความถูกต้องร้อยละ 92.35 ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-11* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-12*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2647               92.3587 %
Incorrectly Classified Instances       219                7.6413 %
Kappa statistic                          0.724
Mean absolute error                      0.1306
Root mean squared error                  0.2647
Relative absolute error                 41.4184 %
Root relative squared error             66.6599 %
Total Number of Instances             2866
```

ภาพที่ 4-11 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 2



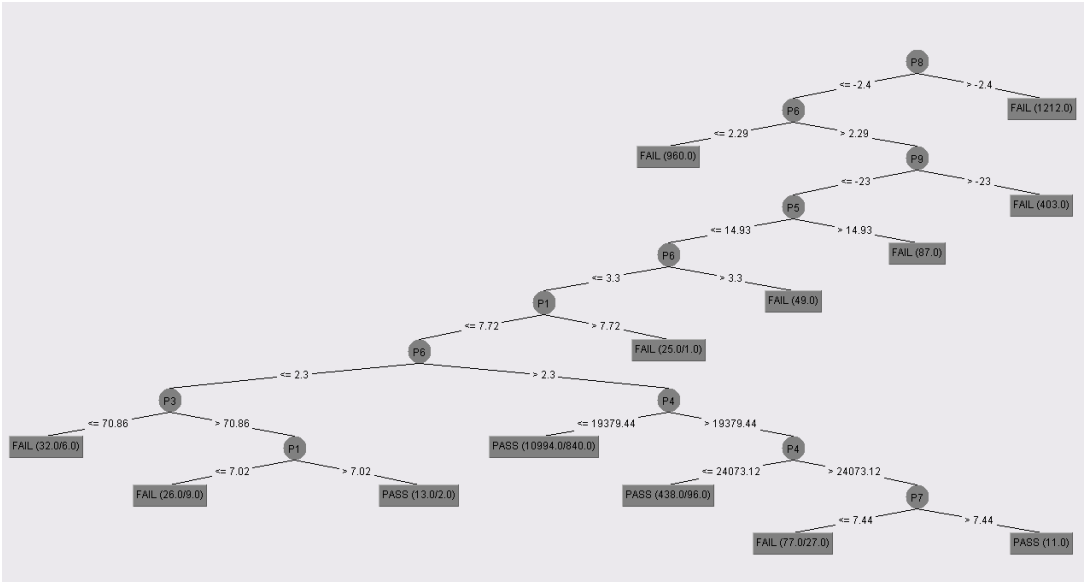
ภาพที่ 4-12 **ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 2

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลทดสอบ จำนวน 14,327 ตัว ด้วย Weka ใช้เวลาการคำนวณ 2.59 วินาทีและให้ความถูกต้องร้อยละ 92.95 ซึ่งแสดงรายละเอียดใน *ภาพที่ 4-13* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-14*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       13317               92.9504 %
Incorrectly Classified Instances      1010                7.0496 %
Kappa statistic                          0.8028
Mean absolute error                      0.1249
Root mean squared error                  0.252
Relative absolute error                 32.2243 %
Root relative squared error             57.2505 %
Total Number of Instances            14327
```

ภาพที่ 4-13 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 2



ภาพที่ 4-14 **ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 2

4.4.3 การปรับพารามิเตอร์

ทำการปรับพารามิเตอร์ทำที่ต้นไม้ตัดสินใจของชุดข้อมูลเรียนรู้ร้อยละ 80 ปรับเฉพาะพารามิเตอร์ที่สามารถทำการปรับได้เท่านั้น ได้แก่ P2-P6, P8 ทำการทดลองปรับพารามิเตอร์ 3 แบบคือบนลงล่าง,ล่างขึ้นบน และ แบบผสม โดยมีเงื่อนไขว่า เมื่อทำการปรั บแล้วไม่สามารถทำให้ชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 มีจำนวนงานดีมากขึ้นจะหยุดทำการปรับทันที

4.4.3.1 การปรับพารามิเตอร์แบบบนลงล่าง

เริ่มต้นที่โหนดรากของต้นไม้ตัดสินใจ ผลจากการปรับแสดงรายละเอียดในตาราง*ที่ 4-9*

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P8 <= -2.4 มีผลทำให้ P6 = 2.895 และ P9 = -37.2987	2,797
P6 > 2.29 มีผลทำให้ P8 = -4.059 และ P9 = -23.4712	2,815
P5<= 14.95 มีผลทำให้ P2 = 323.8589 และ P4 = -20.787	0 (หยุดทำการปรับ)

ตารางที่ 4-9 การปรับข้อมูลของ Product 2 แบบบนลงล่าง

การปรับพารามิเตอร์แบบบนลงล่าง ของ Product 2 สามารถทำได้โดยการปรับ P6 > 2.29

4.4.3.2 การปรับพารามิเตอร์แบบล่างขึ้นบน

เริ่มต้นที่โหนดสุดท้ายของต้นไม้ตัดสินใจ ซึ่งมีด้วยกัน 6 โหนด ผลจากการปรับแต่ละโหนดแสดงรายละเอียดใน*ตารางที่ 4-10, ตารางที่ 4-11, ตารางที่ 4-12, ตารางที่ 4-13, ตารางที่ 4-14 และ ตารางที่ 4-15* ตามลำดับ

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P2 > 311.78 มีผลให้ P4 = -2734.5 และ P5 = 17.451	0 (หยุดทำการปรับ)

ตารางที่ 4-10 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 1

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P7 เป็นพารามิเตอร์ที่ไม่สามารถทำการปรับได้ เลื่อนไปโหนดถัดไป	2,498
P1 เป็นพารามิเตอร์ที่ไม่สามารถทำการปรับได้ เลื่อนไปโหนดถัดไป	2,498
P6 เป็นโหนดที่ไม่สามารถตัดสินว่าเป็น PASS หรือ FAIL ได้ เลื่อนไปโหนดถัดไป	2,498
P2 > 379.67 มีผลให้ P4 = 12826.14 และ P5 = 3.1088	2,531
P8 <= -2.73 มีผลให้ P6 = 2.776 และ P9 = -34.55	2,866
P4 เป็นโหนดที่ไม่สามารถตัดสินว่าเป็น PASS หรือ FAIL ได้ เลื่อนไปโหนดถัดไป	2,866
P4 <= 17820.01 มีผลให้ P2 = 401.82 และ P5 = -1.494	2,866 (หยุดทำการปรับ)

ตารางที่ 4-11 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 2

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P3 <= 72.36	2,498 (หยุดทำการปรับ)

ตารางที่ 4-12 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 3

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P6 > 2.97 มีผลให้ P8 = -2.1616 และ P9 = -39.2851	0 (หยุดทำการปรับ)

ตารางที่ 4-13 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 4

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P7 เป็นพารามิเตอร์ที่ไม่สามารถทำการปรับได้ เลื่อนไปโหนดถัดไป	2,498
P6 เป็นโหนดที่ไม่สามารถตัดสินว่าเป็น PASS หรือ FAIL ได้ เลื่อนไปโหนดถัดไป	2,498
P5 เป็นโหนดที่ไม่สามารถตัดสินว่าเป็น PASS หรือ FAIL ได้ เลื่อนไปโหนดถัดไป	2,498
P6 เป็นโหนดที่ไม่สามารถตัดสินว่าเป็น PASS หรือ FAIL ได้ เลื่อนไปโหนดถัดไป	2,498
P9 <= -23 มีผลให้ P6 = 2.28 และ P8 = -4.12	0 (หยุดทำการปรับ)

ตารางที่ 4-14 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 5

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	2,498
P6 <=3.3 มีผลให้ P8 = -1.269 และ P9 = -46.73	0 (หยุดทำการปรับ)

ตารางที่ 4-15 การปรับข้อมูลของ Product 2 แบบล่างขึ้นบน ที่โหนดที่ 6

การปรับพารามิเตอร์แบบล่างขึ้นบน ของ Product 1 สามารถทำได้โดยการปรับ P2 > 379.67 และ P8 <= -2.73

4.4.3.3 การปรับพารามิเตอร์แบบผสม

ทำการรวมผลที่ได้จากการปรับพารามิเตอร์แบบบนลงล่าง และล่างขึ้นบน จะได้ P2 > 379.67 และ -4.06<= P8 <= -2.73

4.4.3.4 เปรียบเทียบบนลงล่าง,ล่างขึ้นบน และ แบบผสม

ทำการเปรียบเทียบทั้ง 3 วิธีด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20 ซึ่งแสดงรายละเอียดใน*ตารางที่ 4-16* และ ชุดข้อมูลทดสอบซึ่งแสดงรายละเอียดใน*ตารางที่ 4-17*

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	4,498
ปรับแบบบนลงล่าง	2,815
ปรับแบบล่างขึ้นบน	2,866
ปรับแบบผสม	2,866

ตารางที่ 4-16 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 2 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20

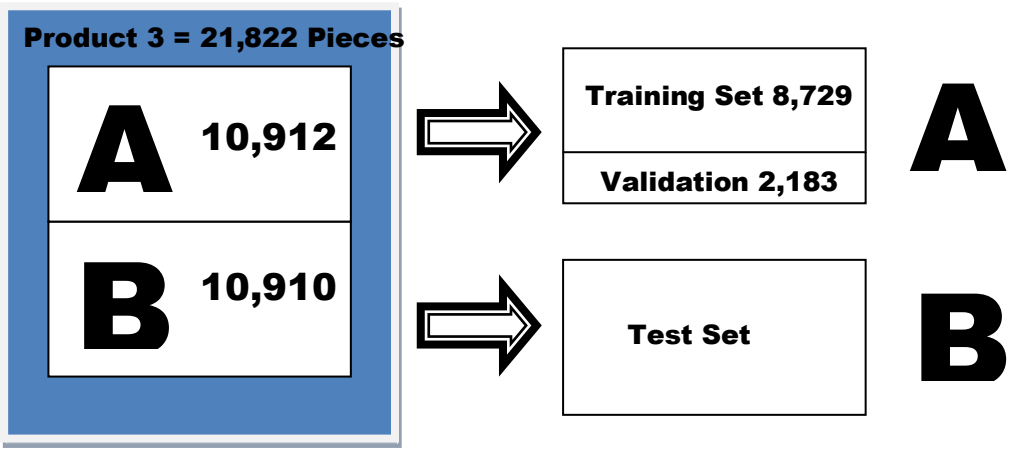
กฎที่ใช้ในการปรับพารามิเตอร์	จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลทดสอบ
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	11,456
ปรับแบบบนลงล่าง	2,069
ปรับแบบล่างขึ้นบน	14,291
ปรับแบบผสม	14,291

ตารางที่ 4-17 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 2 ด้วยชุดข้อมูลทดสอบ

4.5 การทดลองปรับพารามิเตอร์ของ Product 3

4.5.1 การแบ่งข้อมูล

แบ่งข้อมูลออกเป็น 3 ส่วน เช่นเดิมกับการทดลอง Product 1 สัดส่วนการแบ่งข้อมูลแสดงรายละเอียดใน*ภาพที่ 4-15*



ภาพที่ 4-15 สัดส่วนของ Product 3 ในการแบ่งข้อมูลของ Product 3

4.5.2 การสร้างต้นไม้ตัดสินใจ

ในการทดลองนี้ใช้ต้นไม้ตัดสินใจในการหาพารามิเตอร์ที่สำคัญที่มีผลต่อคุณภาพการผลิตฮาร์ดดิสก์ อัลกอริทึมที่นำมาใช้คือ C4.5 ซึ่งทำการสร้างต้นไม้ตัดสินใจด้วยโปรแกรม Weka เลือก Classify ด้วย J48 –C 0.25 M 2 ทำการสร้างต้นไม้ตัดสินใจ 3 ต้นจากชุดข้อมูล

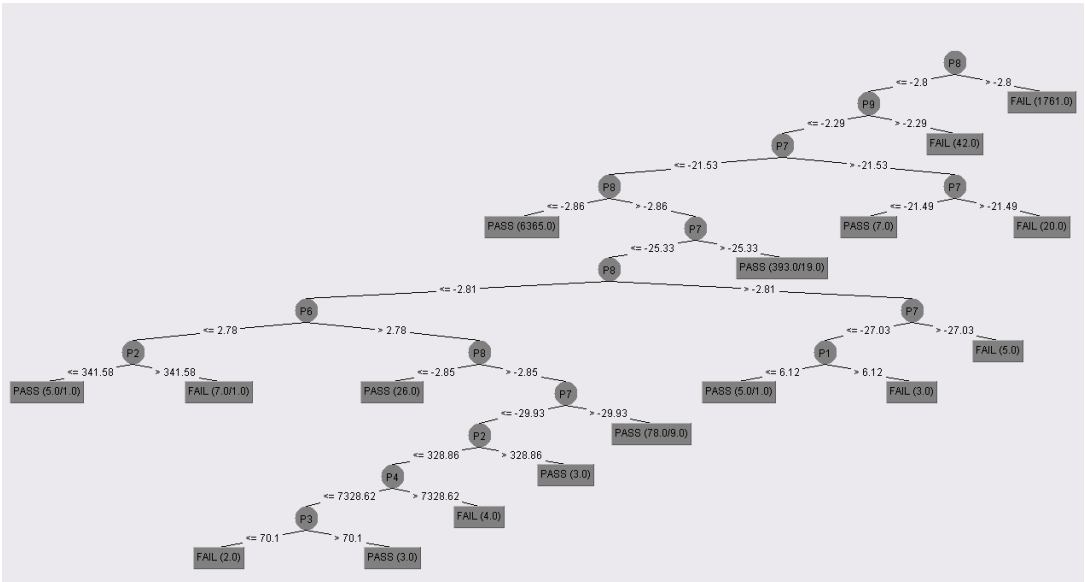
เรียนรู้ร้อยละ 80 และชุดข้อมูลตรวจความสมเหตุสมผลอีกร้อยละ 20 ในส่วนแบ่งแรกและชุดข้อมูลทดสอบในส่วนที่สอง

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลเรียนรู้ร้อยละ 80 จำนวน 8,729 ตัว ด้วย Weka ใช้เวลาการคำนวณ 0.31 วินาทีและให้ความถูกต้องร้อยละ 99.40 ซึ่งแสดงรายละเอียดใน *ภาพที่ 4-16* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4*-17

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        8677               99.4043 %
Incorrectly Classified Instances        52                0.5957 %
Kappa statistic                          0.9822
Mean absolute error                      0.009
Root mean squared error                  0.0739
Relative absolute error                  2.6747 %
Root relative squared error             17.9947 %
Total Number of Instances             8729
```

ภาพที่ 4-16 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 3



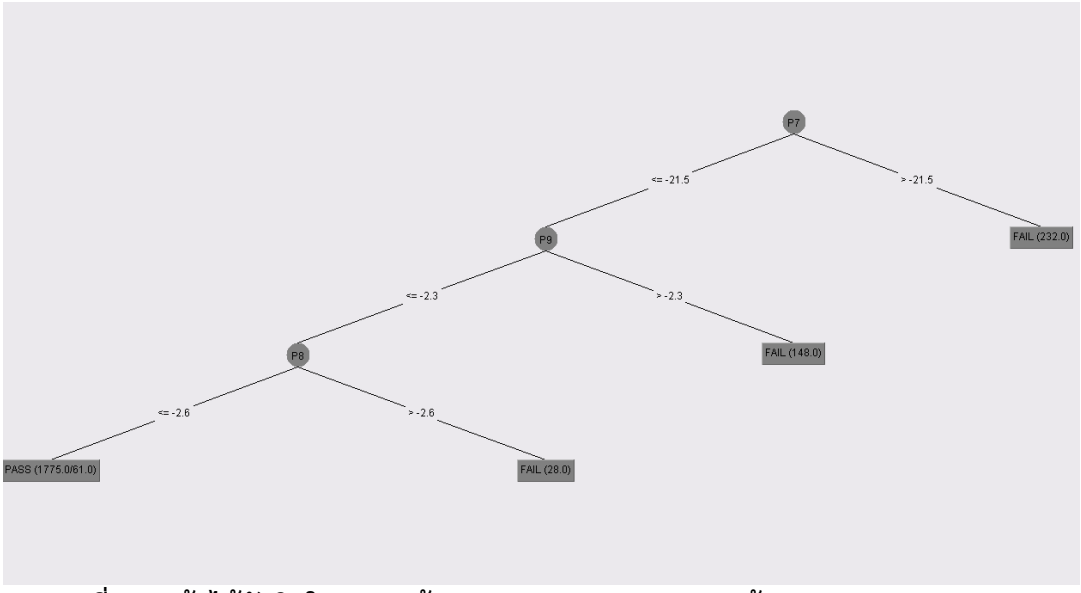
ภาพที่ 4-17 **ต้นไม้ตัดสินใจของชุดข้อมูลการเรียนรู้ร้อยละ** 80 **ของ** Product 3

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 จำนวน 2,183 ตัว ด้วย Weka ใช้เวลาการคำนวณ 0.06 วินาทีและให้ความถูกต้องร้อยละ 96.93 ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-18* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-19*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2116               96.9308 %
Incorrectly Classified Instances        67                3.0692 %
Kappa statistic                          0.9053
Mean absolute error                      0.0519
Root mean squared error                  0.1684
Relative absolute error                 15.373  %
Root relative squared error             40.9926 %
Total Number of Instances             2183
```

ภาพที่ 4-18 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 3



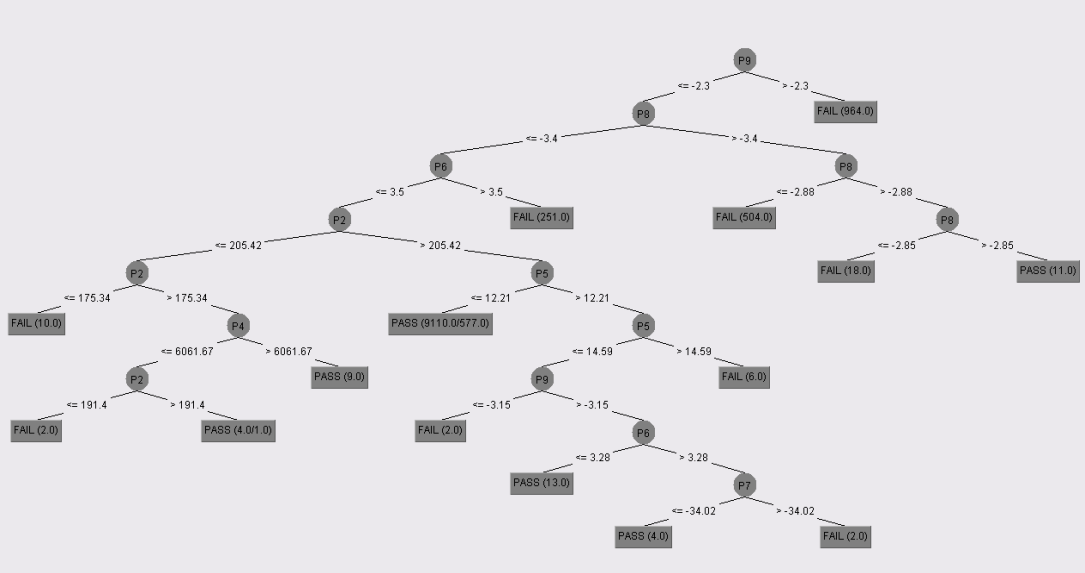
ภาพที่ 4-19 **ต้นไม้ตัดสินใจของชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ** 20 **ของ** Product 3

การสร้างต้นไม้ตัดสินใจจากชุดข้อมูลทดสอบ จำนวน 10,910 ตัว ด้วย Weka ใช้เวลาการคำนวณ 1.19 วินาทีและให้ความถูกต้องร้อยละ 94.56 ซึ่งแสดงรายละเอียดใน *ภาพที่ 4-20* และภาพต้นไม้ตัดสินใจที่ได้ซึ่งแสดงรายละเอียดใน*ภาพที่ 4-21*

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       10316               94.5555 %
Incorrectly Classified Instances       594                5.4445 %
Kappa statistic                          0.8226
Mean absolute error                      0.1
Root mean squared error                  0.2255
Relative absolute error                 29.7041 %
Root relative squared error             54.9742 %
Total Number of Instances            10910
```

ภาพที่ 4-20 **รายงานต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 3



ภาพที่ 4-21 **ต้นไม้ตัดสินใจของชุดข้อมูลทดสอบ ของ** Product 3

4.5.3 การปรับพารามิเตอร์

ทำการปรับพารามิเตอร์ทำที่ต้นไม้ตัดสินใจของชุดข้อมูลเรียนรู้ร้อยละ 80 ปรับเฉพาะพารามิเตอร์ที่สามารถทำการปรับได้เท่านั้น ได้แก่ P1-P3, P6-P9 ทำการทดลองปรับพารามิเตอร์ 3 แบบคือบนลงล่าง ล่างขึ้นบน และ แบบผสม โดยมีเงื่อนไขว่า เมื่อทำการปรับแล้วไม่สามารถทำให้ชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 มีจำนวนงานดีมากขึ้นจะหยุดทำการปรับทันที

4.5.3.1 การปรับพารามิเตอร์แบบบนลงล่าง

เริ่มต้นที่โหนดรากของต้นไม้ตัดสินใจ ผลจากการปรับแสดงรายละเอียดใน *ตารางที่ 4-18*

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
P8 <= -2.8 มีผลทำให้ P6 = 2.9, P7 = -24.12 และ P9 = -3.47	2,183 (หยุดทำการปรับ)

ตารางที่ 4-18 การปรับข้อมูลของ Product 3 แบบบนลงล่าง

การปรับพารามิเตอร์แบบบนลงล่าง ของ Product 3 สามารถทำได้โดยการปรับ P8 <= -2.8

4.5.3.2 การปรับพารามิเตอร์แบบล่างขึ้นบน

เริ่มต้นที่โหนดสุดท้ายของต้นไม้ตัดสินใจ ซึ่งมีด้วยกัน 4 โหนด ผลจากการปรับแต่ละโหนดแสดงรายละเอียดใน *ตารางที่ 4-19, ตารางที่ 4-20, ตารางที่ 4-21 และ ตารางที่ 4-22* ตามลำดับ

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
P3 > 70.1	1,775 (หยุดทำการปรับ)

ตารางที่ 4-19 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 1

กฎที่ใช้ในการปรับพารามิเตอร์	**จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20**
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
P2 <= 341.58 มีผลให้ P4 = 15780.01 และ P5 = -1.21344	1,775 (หยุดทำการปรับ)

ตารางที่ 4-20 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 2

กฎที่ใช้ในการปรับพารามิเตอร์	จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
P1 <= 6.12	1,775 (หยุดทำการปรับ)

ตารางที่ 4-21 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 3

กฎที่ใช้ในการปรับพารามิเตอร์	จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
P7 <= -21.49 มีผลให้ P6 = 2.8, P8 = -2.4 และ P9 = -3.86	1,775 (หยุดทำการปรับ)

ตารางที่ 4-22 การปรับข้อมูลของ Product 3 แบบล่างขึ้นบน ที่โหนดที่ 4

การปรับพารามิเตอร์แบบล่างขึ้นบน ของ Product 3 ไม่สามารถทำการปรับได้

4.5.3.3 การปรับพารามิเตอร์แบบผสม

ทำการรวมผลที่ได้จากการปรับพารามิเตอร์แบบ บนลงล่าง และล่างขึ้นบน แต่เนื่องจาก การปรับพารามิเตอร์แบบ ล่างขึ้นบน ไม่สามารถทำการปรับค่าได้ จึงทำให้ การปรับพารามิเตอร์แบบผสม มีค่าเท่ากับ การปรับพารามิเตอร์แบบบนลงล่าง คือ P8 <= -2.8

4.5.3.4 เปรียบเทียบบนลงล่าง ล่างขึ้นบน และแบบผสม

ทำการเปรียบเทียบทั้ง 3 วิธีด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20 ซึ่งแสดงรายละเอียดใน*ตารางที่ 4-23* และ ชุดข้อมูลทดสอบซึ่งแสดงรายละเอียดใน*ตารางที่ 4-24*

กฎที่ใช้ในการปรับพารามิเตอร์	จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	1,775
ปรับแบบบนลงล่าง	2,183
ปรับแบบล่างขึ้นบน	1,775
ปรับแบบผสม	2,183

ตารางที่ 4-23 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 3 ด้วยชุดข้อมูลตรวจสอบความสมเหตุสมผลร้อยละ 20

กฎที่ใช้ในการปรับพารามิเตอร์	จำนวนชิ้นงานที่มีคุณภาพในชุดข้อมูลทดสอบ
เริ่มต้นไม่มีการปรับพารามิเตอร์ใดๆ	9,151
ปรับแบบบนลงล่าง	10,910
ปรับแบบล่างขึ้นบน	9,151
ปรับแบบผสม	10,910

ตารางที่ 4-24 ตารางเปรียบเทียบวิธีการปรับพารามิเตอร์ทั้ง 3 วิธี ของ Product 3 ด้วยชุดข้อมูลทดสอบ

4.6 วิเคราะห์ผลการทดลอง

จากผลการทดลอง ด้วยตัวอย่างผลิตภัณฑ์ฮาร์ดดิสก์ 3 ชนิด ด้วย วิธีการปรับพารามิเตอร์ที่นำเสนอ 3 วิธีคือ การปรับพารามิเตอร์แบบ บนลงล่าง และล่างขึ้นบน ซึ่งมีวิธีการปรับพารามิเตอร์ที่คล้ายกัน คือ อ้างอิงชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 หากให้จำนวนงานดีมากขึ้นจะทำการปรับพารามิเตอร์ต่อ แตกต่างกันที่จุดเริ่มต้นในการปรับ ซึ่ง บนลงล่าง เริ่มต้นที่โหนดราก และ ล่างขึ้นบน เริ่มต้นที่โหนดของใบ จากผลการทดลอง *ตารางที่ 4.7 และ ตารางที่ 4.8* ของผลิตภัณฑ์ชนิดที่ 1 พบว่าเมื่อทำการปรับทั้งสองวิธี ต่างทำให้จำนวนของงานดีเพิ่มมากขึ้นและเมื่อนำผลที่ได้จากทั้งสองวิธีมารวมกัน หรือ เรียกวิธีนี้ว่า ปรับ แบบผสม จากผลการทดลองพบว่า สามารถให้จำนวนงานดีเพิ่มขึ้นมากที่สุดเมื่อเปรียบเทียบกับวิธีทั้งหมด *ตารางที่ 4.16* และ *ตารางที่ 4.17* ของผลิตภัณฑ์ชนิดที่ 2 พบว่าเมื่อทำการปรับทั้งสองวิธี แบบ ล่างขึ้นบน

ทำให้จำนวนของงานดีเพิ่มมากขึ้นมากกว่าและเมื่อนำผลที่ได้จากทั้งสองวิธีมารวมกัน หรือ เรียกวิธีนี้ว่า ปรับ แบบผสม จะได้ผลเท่ากับการการปรับค่าที่ดีที่สุด คือ แบบ ล่างขึ้นบน เมื่อทำการทดลองกับชุดข้อมูลทดสอบพบว่าการปรับค่าแบบผสม และล่างขึ้นบน ทำให้จำนวนของงานดีเพิ่มมากขึ้นมากที่สุด *ตารางที่ 4.23* และ *ตารางที่ 4.24* ของผลิตภัณฑ์ชนิดที่ 3 พบว่าเมื่อทำการปรับทั้งสองวิธี แบบบนลงล่าง ทำให้จำนวนของงานดีเพิ่มมากขึ้นมากกว่าและเมื่อนำผลที่ได้จากทั้งสองวิธีมารวมกัน หรือ เรียกวิธีนี้ว่า ปรับแบบผสม จะได้ผลเท่ากับการการปรับค่าที่ดีที่สุด คือ แบบบนลงล่าง เมื่อทำการทดลองกับชุดข้อมูลทดสอบพบว่าการปรับค่าแบบผสม และบนลงล่าง ทำให้จำนวนของงานดีเพิ่มมากขึ้นมากที่สุดซึ่งเป็นที่น่าสังเกตว่า การปรับค่า แบบผสม นั้น สามารถให้จำนวนของงานดีเพิ่มมากขึ้นมากที่สุดไม่ว่าจะทดสอบกับชุดข้อมูลของผลิตภัณฑ์ฮาร์ดดิสก์ไหน

4.7 สรุปผลการทดลอง

จากการทดลองการปรับพารามิเตอร์ทั้ง 3 วิธีกับชุดข้อมูลตรวจความสมเหตุสมผลร้อยละ 20 และชุดข้อมูลทดสอบพบว่า วิธีการปรับพารามิเตอร์แบบผสม สามารถ ให้จำนวนงานดีเพิ่มมากขึ้นมากที่สุด เนื่องจากเป็นการนำผลการปรับพารามิเตอร์ด้วยวิธี บนลงล่าง และล่างขึ้นบน มารวมกันจึงทำให้การปรับพารามิเตอร์นั้นครอบคลุมทุกกรณีที่เกิดงานเสียมากที่สุด นั่นคือ ต้องทำการทดลองทั้ง2 แบบ คือบนลงล่าง กับล่างขึ้นบน เพื่อนำไปหาการปรับค่าแบบผสม ที่จะนำไปสร้างแบบจำลองการพัฒนาคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ต่อไป

บทที่ 5

สรุปผลการวิจัย และแนวทางการพัฒนาต่อ

สรุปผลการวิจัย

งานวิจัยนี้เสนอการสร้างแบบจำลองการพัฒนาคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์เพื่อเป็นแนวทางในการแก้ปัญหาการผลิตในโรงงาน โดยใช้วิธีการปรับพารามิเตอร์จากต้นไม้ตัดสินใจซึ่งทำการทดลอง 3 วิธี คือบนลงล่าง ล่างขึ้นบน และแบบผสม เพื่อหาวิธีที่ดีที่สุด การทดลองแบบบนลงล่าง นั้นทำการปรับพารามิเตอร์เริ่มตั้งแต่โหนดรากลงไป จนกระทั่งปรับแล้วไม่ทำให้จำนวนงานดีเพิ่มขึ้น ส่วนการทดลองแบบล่างขึ้นบน นั้นทำการปรับพารามิเตอร์ตั้งแต่โหนดใบขึ้นไปเรื่อยๆ จนกระทั่งปรับแล้วไม่ทำให้จำนวนงานดีเพิ่มขึ้นเช่นกัน และการทดลอง แบบผสมเป็นการรวมผลที่ได้จากการปรับแบบบนลงล่าง และล่างขึ้นบน ซึ่งผลที่ได้จากการทดลองสามารถสรุปได้ว่า วิธีแบบผสม เป็นวิธีที่ดีที่สุด เพราะเกิดจากการนำวิธีการปรับพารามิเตอร์ที่ได้จากทั้งสองวิธีมาทำการปรับพารามิเตอร์พร้อมกันซึ่งจะสามารถทำให้เพิ่มจำนวนงานดีได้มากที่สุด ดังนั้นสามารถสรุปได้ว่า การปรับพารามิเตอร์แบบผสม เหมาะสมที่สุดที่จะนำไปสร้างเป็นแบบจำลองในการพัฒนาคุณภาพการผลิตฮาร์ดดิสก์

แนวทางในการพัฒนาต่อ

วิธีที่นำเสนอนี้เป็นวิธีหนึ่งที่ใช้ในการปรับพารามิเตอร์จากต้นไม้ตัดสินใจ ซึ่งการปรับนั้นมุ่งสนใจไปที่จำนวนงานดีที่เพิ่มขึ้น แต่ยังไม่คลอบคลุมถึงจำนวนเวลาที่ใช้ในการคำนวณและระยะทางที่สั้นที่สุด ที่จะแก้ไขปัญหานี้ โดยคำนึงถึงต้นทุนการผลิต ซึ่งสามารถใช้เป็นแนวทางในการศึกษาพัฒนาต่อเพื่อให้วิธีการนี้มีประสิทธิภาพที่ดีและครอบคลุมต่อทุกการใช้งานมากขึ้น

รายการอ้างอิง

[1] The Board of Investment of Thailand (BOI). Annual Report 2010. Bangkok: THAILAND BOARD OF INVESTMENT, 2010.

[2] Mitchell, T.M. Machine Learning. New York: McGraw-hill, 1997.

[3] Kijsirikul, B. Data Mining Algorithms. Chulalongkorn University: the Joint Government and Private Sectors, 2004.

[4] Remco, R., Bouckaert, E.F., Mark, A.H., Geoffrey, H., Bernhard, P., Peter, R., and Lan, H. W. WEKA-experiences with a java open-source project. Journal of Machine Learning Research. 11(2010): 2533-2541.

[5] Quinlan, J. R. Induction of decision trees. Mach. Learn.1(1986): 81-106.

[6] Quinlan, J. R. C4.5: Programs for Machine Learning. San Francisco: Morgan Kaufmann Publishers, 1993.

[7] Breiman, L., Friedman, J., Olshen, R., and Stone, C. Classification and Regression Trees. Belmont :Wadsworth, 1984.

[8] Kass, G. An exploratory technique for investigating large quantities of categorical data. Applied Statistics. Appl.Statist. 29(1980): 119-127.

[9] Taetragool, U., and Achalakul, T. Applying Decision Tree in Fault Pattern Analysis for HGA ManuFacturing., Int. Conf. on Complex, Intelligent and Software Intensive Systems, pp.83-89. Japan:IEEE, 2009.

[10] Huang, Y.T., Cheng, F.T., and Hung. M.H. Developing a Product Quality Fault Detection Scheme., IEEE Int. Conf. on Robotics and Automation, pp.927–932. Japan:IEEE, 2009.

[11] Yamwong, W.,Kaotien, J., and Achalakul, T. The Sampling-based Sensitivity Analysis Model for Yield Improvement in HDD Manufacturing., Int. Conf. on Complex, Intelligent and Software Intensive Systems, pp. 1211–1216. Japan:IEEE, 2009.

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวอโณทัย ศิลเทพาเวทย์ เกิดเมื่อวันที่ 8 มกราคม 2527 ในกรุงเทพมหานคร จบการศึกษาชั้นมัธยมศึกษาที่โรงเรียนเตรียมอุดมศึกษา กรุงเทพมหานคร จบการศึกษาในระดับปริญญาบัณฑิตจากภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล เมื่อปีการศึกษา 2549 และในระดับปริญญามหาบัณฑิตจากภาควิชาวิศวกรรมศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2554