

วิธีวิวัตติกเพื่อปรับปรุงต้นไม้ตัดสินใจสำหรับยอมรับผลิตภัณฑในการผลิตฮาร์ดดิสก์

A Heuristic Approach to Improve The Decision Tree of Product Acceptance in Hard Drive Manufacturing

บทคัดย่อ

บทความนี้เสนอการสร้างแบบจำลองเพื่อพัฒนาคุณภาพการผลิตฮาร์ดดิสก์ด้วยเทคนิคต้นไม้ตัดสินใจ แบบจำลองนี้ช่วยบอกแนวทางการปรับพารามิเตอร์ในโรงงานให้ได้จำนวนงานที่ตีมากที่สุด วิธีการแบ่งออกเป็นสองส่วน ส่วนแรกกระบวนการพารามิเตอร์ที่สำคัญในโรงงานด้วยต้นไม้ตัดสินใจ ส่วนที่สองเป็นการปรับพารามิเตอร์ ในงานวิจัยนี้ทำการทดลองปรับพารามิเตอร์ 2 วิธี คือ ปรับจากพารามิเตอร์ที่โรงงานเสียมากที่สุดก่อน และ ปรับจากพารามิเตอร์ที่โรงงานเสียน้อยที่สุดก่อน จากนั้นทำการเปรียบเทียบ 2 วิธีนี้และนำวิธีที่ดีที่สุดที่สามารถเพิ่มจำนวนงานตีได้มากที่สุดไปสร้างเป็นแบบจำลองในการพัฒนาคุณภาพของฮาร์ดดิสก์

คำสำคัญ: ต้นไม้ตัดสินใจ การผลิตฮาร์ดดิสก์ การพัฒนาคุณภาพของผลิตภัณฑ

Abstract

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 maximum pass-pieces. 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. Furthermore, two approaches are illustrated to identify the adjustment

methods, i.e. the Maximum-Fail First search, the Minimum-Fail First search then the best method which gains more quantity of pass-pieces will be selected.

Keyword: Decision Tree, Hard drive Manufacturing, Improving Quality of Product.

1. บทนำ

อุตสาหกรรมอิเล็กทรอนิกส์และเครื่องใช้ไฟฟ้าถือเป็นอุตสาหกรรมที่สำคัญในประเทศไทยมียอดการส่งออกสูงที่สุดซึ่งอุตสาหกรรมอิเล็กทรอนิกส์ ได้แก่ ฮาร์ดดิสก์ [1] กระบวนการผลิตฮาร์ดดิสก์แบ่งออกเป็นหลายขั้นตอน ซึ่งในแต่ละขั้นตอนจะมีการทดสอบพารามิเตอร์ของผลิตภัณฑเพื่อมาจำแนกประเภทว่าเป็นงานที่ผลิตได้คุณภาพดีหรือเสีย งานวิจัยนี้มุ่งเน้นที่จะควบคุมกระบวนการผลิตโดยสนใจเฉพาะการปรับพารามิเตอร์ที่สามารถควบคุมได้ในโรงงาน เพื่อเพิ่มจำนวนงานที่มีคุณภาพดีมากขึ้น ลดของเสียในโรงงาน ด้วยการสร้างแบบจำลองเพื่อช่วยในการควบคุมตัวแปรในโรงงาน ซึ่งแบ่งออกเป็น 2 ส่วน ส่วนแรก นำพารามิเตอร์ทั้งหมดของการผลิตมาสร้างต้นไม้ตัดสินใจเพื่อหาพารามิเตอร์ที่มีผลต่อคุณภาพของผลิตภัณฑโดยตรง ส่วนที่สองทำการจำแนกตัวแปรออกเป็น 2 ประเภทคือ พารามิเตอร์ที่สามารถควบคุมได้ และ พารามิเตอร์ที่ไม่สามารถควบคุมได้ งานวิจัยนี้มุ่งหมายที่จะทำการปรับค่าเฉพาะพารามิเตอร์ที่สามารถทำการควบคุมได้เท่านั้น โดยทำการทดลองปรับจากต้นไม้ตัดสินใจ 2 วิธี คือ ปรับจากพารามิเตอร์ที่โรงงานเสียมากที่สุดก่อน กล่าวคือเริ่มต้นทำการปรับที่พารามิเตอร์ที่มีผลทำให้เกิดจำนวนงานเสียมากที่สุด และ ปรับจากพารามิเตอร์ที่โรงงานเสียน้อยที่สุดก่อน กล่าวคือเริ่มต้นทำการปรับที่พารามิเตอร์ที่มีผลทำให้เกิดจำนวนงานเสีย

น้อยที่สุด จากนั้นทำการเปรียบเทียบวิธีการทั้งสองเพื่อหาวิธีที่สามารถเพิ่มจำนวนงานได้ดีมากที่สุดมาสร้างแบบจำลองการควบคุมพารามิเตอร์ในกระบวนการผลิตฮาร์ดดิสก์

2. งานวิจัยที่เกี่ยวข้อง

2.1 ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีหนึ่งในการเรียนรู้ของเครื่อง (Machine Learning) ที่นิยมใช้กันอย่างแพร่หลาย [2] ใช้สำหรับการจำแนกข้อมูลตามคุณลักษณะที่สำคัญแสดงให้อยู่ในรูปแบบเงื่อนไข IF-THEN ซึ่งสามารถทำความเข้าใจได้ง่าย ต้นไม้ตัดสินใจมีประสิทธิภาพการประมวลผลได้เร็วเมื่อเทียบกับอัลกอริทึมการจำแนกประเภทอื่นๆ [3] จึงเหมาะที่จะใช้ในการประมวลผลข้อมูลของโรงงานซึ่งมีขนาดใหญ่

2.2 โปรแกรม WEKA

เป็นเครื่องมือการทำภาษาเครื่องถูกพัฒนาจาก The University of Waikato in New Zealand ปี 1997 ซึ่งประกอบไปด้วยอัลกอริทึมต่างๆ ในภาษาเครื่องและเครื่องมือสำหรับทำเหมืองข้อมูล [4] จึงเหมาะที่จะนำมาใช้ในการทดลองสร้างต้นไม้ตัดสินใจในงานวิจัยนี้

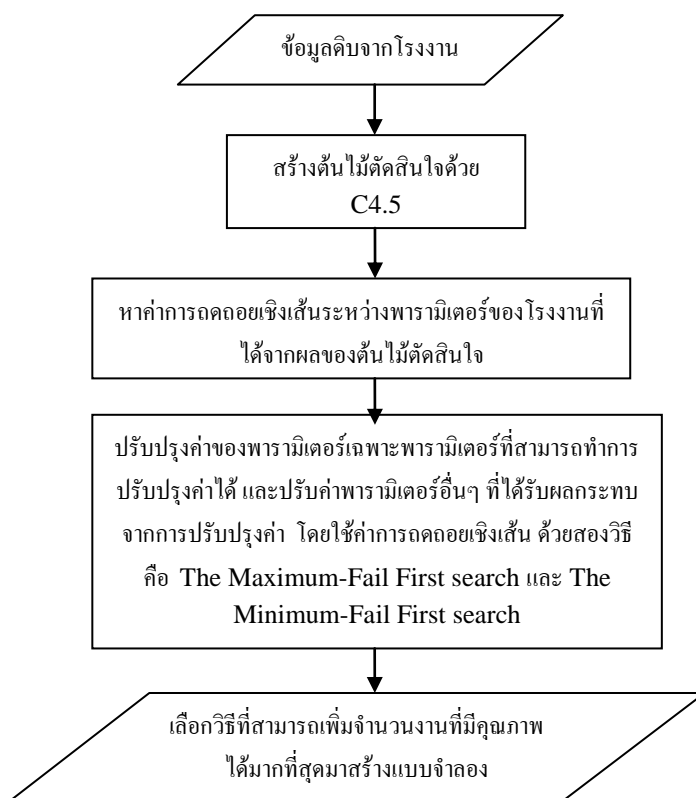
2.3 งานวิจัยที่เกี่ยวกับต้นไม้ตัดสินใจและฮาร์ดดิสก์

อัลกอริทึมที่ใช้จำแนกความดีของคุณลักษณะในต้นไม้ตัดสินใจมีมากมายได้แก่ ID3 [5], C4.5 [6], CART [7] และ CHAID [8] มีงานวิจัยเกี่ยวกับ C4.5 เพื่อใช้ในการหารูปแบบของงานเสียในการผลิต HGA ซึ่งเป็นส่วนประกอบของฮาร์ดดิสก์ [9] ได้ทำการเปรียบเทียบ C4.5 กับอัลกอริทึมอื่นๆ พบว่า C4.5 ถูกพัฒนามาจาก ID3 เพิ่มประสิทธิภาพทางการคำนวณให้ดีขึ้น หลีกเลี่ยงสถานะข้อมูลที่เหมาะสมไป ซึ่งมักเกิดกับข้อมูลที่มีขนาดใหญ่เช่นข้อมูลโรงงานอุตสาหกรรม ใน CART ใช้จำแนกข้อมูลที่เป็นค่าต่อเนื่อง จึงไม่เหมาะกับงานวิจัยนี้จะใช้จำแนกข้อมูลออกเป็นคุณลักษณะ ส่วน CHAID ไม่มีส่วนของการตัดเล็มต้นไม้ หากใช้กับข้อมูลของโรงงานจะทำให้เกิดสถานะข้อมูลที่เหมาะสมไป งานวิจัยนี้จึงเลือกที่จะทำการพัฒนาด้วย C4.5 เพื่อใช้ระบุพารามิเตอร์ที่มีผลโดยตรงต่อคุณภาพของผลิตภัณฑ์ ในส่วนแรก งานวิจัย [10] เกี่ยวข้องกับการตรวจจับของเสียในการผลิตสารกึ่งตัวนำโดยใช้ CART ในการสร้างแบบจำลองเพื่อระบุความสัมพันธ์ระหว่าง

พารามิเตอร์กับกระบวนการผลิต โดยประยุกต์ใช้หลักการต้นทุนต่ำสุดสำหรับตัดเล็มต้นไม้เพื่อหลีกเลี่ยงสถานะข้อมูลที่เหมาะสมไป งานวิจัย [11] เสนอเครื่องมือพัฒนากระบวนการผลิตฮาร์ดดิสก์ โดยทำการหาพารามิเตอร์และคุณลักษณะที่มีผลกระทบต่อคุณภาพของผลิตภัณฑ์ด้วยต้นไม้ตัดสินใจและทำการพัฒนาคุณภาพด้วยการปรับค่าเฉลี่ยของพารามิเตอร์เพื่อให้ได้คุณภาพของผลิตภัณฑ์ที่สูงขึ้น

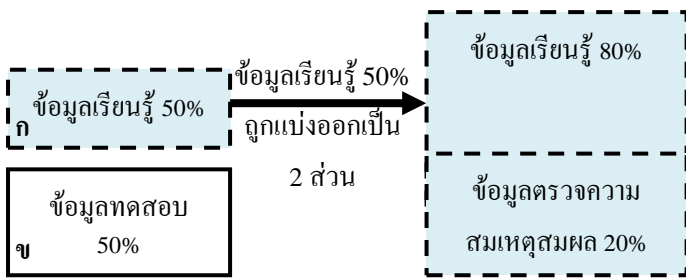
3. แนวคิดและวิธีการดำเนินงาน

จุดมุ่งหมายของงานวิจัยนี้เพื่อต้องการออกแบบการควบคุมพารามิเตอร์ของโรงงานซึ่งมีผลโดยตรงต่อคุณภาพของผลิตภัณฑ์เพื่อให้ได้จำนวนของผลิตภัณฑ์ที่ผ่านคุณภาพมากขึ้น โดยอะแกรมของแนวคิดได้ถูกแสดงไว้ที่ภาพที่ 1 เฉพาะพารามิเตอร์ที่สามารถทำการปรับได้เท่านั้นที่จะทำการปรับปรุงเพื่อลดจำนวนงานเสียลงและเพิ่มจำนวนงานที่มีคุณภาพมากขึ้น อย่างไรก็ตามการปรับค่าพารามิเตอร์อาจส่งผลกระทบต่อพารามิเตอร์อื่นๆ ดังนั้น ในการทดลองจึงใช้ค่าการถดถอยเชิงเส้นในการหาความสัมพันธ์ระหว่างพารามิเตอร์ในโรงงานเพื่อใช้ในการหาค่าพารามิเตอร์เมื่อพารามิเตอร์ที่มีความสัมพันธ์นั้นมีการปรับค่าจากการทดลอง



ภาพที่ 1: ไดอะแกรมแนวคิด

เนื่องจากในการทดลองไม่สามารถทำการปรับปรุงค่าพารามิเตอร์ในกระบวนการผลิตได้จริง เพราะอาจส่งผลให้เกิดความเสียหายในโรงงาน การทดลองจึงทำการแบ่งข้อมูลโรงงานออกเป็นสองส่วนเท่าๆ กัน โดยข้อมูลในส่วนแรกจะใช้เป็นชุดข้อมูลเรียนรู้ และส่วนที่สองเป็นชุดข้อมูลทดสอบ ในส่วนของชุดข้อมูลเรียนรู้นั้นถูกแบ่งออกอีก 20 เปอร์เซ็นต์ เพื่อเป็นชุดข้อมูลตรวจสอบความสมเหตุสมผลซึ่งจะใช้ในการตรวจว่าเมื่อไหร่ควรทำการหยุดทำการปรับปรุงค่าพารามิเตอร์ในต้นไม้ตัดสินใจ ซึ่งสัดส่วนของข้อมูลโรงงานทั้งหมดได้ถูกแสดงไว้ที่ภาพที่ 2



ภาพที่ 2: สัดส่วนของข้อมูลโรงงานทั้งหมด

ข้อมูลเรียนรู้ 80% ถูกใช้ในการสร้างต้นไม้ตัดสินใจที่แสดงพารามิเตอร์ที่สามารถปรับค่าให้จำนวนงานที่มีคุณภาพเพิ่มขึ้น ซึ่งจากต้นไม้ตัดสินใจที่ได้สามารถทำการปรับค่าพารามิเตอร์ได้ 2 วิธี คือ ปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อน และ ปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อน ซึ่งทั้งสามวิธีได้อธิบายไว้ที่ ภาพที่ 3, ภาพที่ 4 และ นอกเหนือจากนั้นวิธีการที่ใช้ในการปรับค่าพารามิเตอร์ที่ได้รับผลกระทบเนื่องจากค่าพารามิเตอร์ที่กำลังทำการถูกปรับ ในขณะที่มีความสัมพันธ์กัน ได้ทำการอธิบายไว้ที่ภาพที่ 5

MAXIMUM FAIL SEARCHING $[T, V, T']$
 Input: T , a set of parameters of a decision tree order by maximum fail quantity to minimum fail quantity
 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: $i \leftarrow 0$
 1 WHILE ($Accuracy$ is increasing)
 2 IF T_i is controllable parameter

3 Add a pair of parameter and its value in T_i to L .
 4 $Accuracy \leftarrow EVALUATE[V, T', L]$
 5 $i \leftarrow i + 1$
 6 RETURN L

ภาพที่ 3: THE MAXIMUM-FAIL FIRST SEARCH อัลกอริทึม

MINIMUM FAIL SEARCHING $[T, V, T']$
 Similar to MAXIMUM FAIL SEARCHING except T is a set of parameters of a decision tree order by minimum fail quantity to maximum fail quantity

ภาพที่ 4: THE MINIMUM-FAIL FIRST SEARCH อัลกอริทึม

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

ภาพที่ 5: อัลกอริทึมการปรับค่าการลดหย่อนเชิงเส้น

3. ผลการทดลอง

ข้อมูลที่น่ามาใช้ในการทดลองเป็นข้อมูลของการผลิตฮาร์ดดิสก์ของผลิตภัณฑ์ชนิดหนึ่งใน 1 วัน ซึ่งมีจำนวน 64,887 ชิ้น ประกอบไปด้วย 12 พารามิเตอร์ ซึ่งแสดงใน ตารางที่ 1 ข้อมูลถูกแบ่งเป็นสองส่วนตามวิธีการข้างต้น กลุ่มแรกเป็นข้อมูลเรียนรู้(ก) จำนวน 32,443 ชิ้น และ อีกกลุ่มเป็นข้อมูลทดสอบ(ข) 32,444 ชิ้น

ตารางที่ 1: ตัวอย่างของข้อมูล โรงงานผลิตฮาร์ดดิสก์ซึ่งแสดงรายละเอียดของพารามิเตอร์และคุณลักษณะ

ชื่อพารามิเตอร์	รายละเอียด
P1, P2, P3, P4, P5, P6	พารามิเตอร์ที่สามารถควบคุมได้
P7	พารามิเตอร์ที่มีความสัมพันธ์กับ P3
P8	พารามิเตอร์ที่มีความสัมพันธ์กับ P2, P4
P9, P10, P11	พารามิเตอร์ที่ไม่สามารถควบคุมได้

การปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อนเริ่มต้นทำการปรับตามลำดับที่มากไปหาน้อยตามตารางที่ 2 จนกระทั่ง

เมื่อค่าที่ปรับกับกลุ่มข้อมูลทดสอบ 20 เปอร์เซ็นต์จากกลุ่มข้อมูลเรียนรู้ไม่สามารถเพิ่มจำนวนงานดีขึ้นได้

ตารางที่ 2: ตัวอย่างของข้อมูลโรงงานผลิตรถยนต์ซึ่งแสดงผลของการปรับค่าพารามิเตอร์ด้วยวิธีการปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อน

ลำดับ	ชื่อพารามิเตอร์	จำนวนงานเสีย	จำนวนงานดี
1	P5 (1)	4311	4777
2	P6	1235	5097
3	P10	146	พารามิเตอร์ที่ไม่สามารถควบคุมได้ (ผ่าน)
4	P9 (1)	102	พารามิเตอร์ที่ไม่สามารถควบคุมได้ (ผ่าน)
5	P8 (1)	70	0 (หยุด)
6	P11	55	-
7	P9 (2)	37	-
8	P5 (2)	9	-
9	P1	7	-
10	P9 (3)	6	-
11	P8 (2)	5	-
12	P9 (4)	3	-

จากตารางที่ 2 พบว่า การปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อนสามารถปรับได้สองพารามิเตอร์ คือ P5 และ P6

การปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อนทดลองคล้ายกับการปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อนแต่เริ่มต้นทำการปรับตามลำดับที่น้อยไปหามากตามตารางที่ 3 จนกระทั่งเมื่อค่าที่ปรับกับกลุ่มข้อมูลทดสอบ 20 เปอร์เซ็นต์จากกลุ่มข้อมูลเรียนรู้ไม่สามารถเพิ่มจำนวนงานดีขึ้นได้

ตารางที่ 3: ตัวอย่างของข้อมูล โรงงานผลิตรถยนต์ซึ่งแสดงผลของการปรับค่าพารามิเตอร์ด้วยวิธีการปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อน

ลำดับ	ชื่อพารามิเตอร์	จำนวนงานเสีย	จำนวนงานดี
1	P9 (4)	3	พารามิเตอร์ที่ไม่สามารถควบคุมได้ (ผ่าน)

2	P8 (2)	5	4637
3	P9 (3)	6	พารามิเตอร์ที่ไม่สามารถควบคุมได้ (ผ่าน)
4	P1	7	4637(หยุด)
5	P5 (2)	9	-
6	P9 (2)	37	-
7	P11	55	-
8	P8 (1)	70	-
9	P9 (1)	102	-
10	P10	146	-
11	P6	1235	-
12	P5 (1)	4311	-

จากตารางที่ 3 พบว่า การปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อนสามารถปรับได้แค่พารามิเตอร์เดียว คือ P8 ซึ่งมีความสัมพันธ์กับพารามิเตอร์ P2 และ P4 จึงต้องทำการปรับค่าด้วยสมการการถดถอยเชิงเส้น

เมื่อทำการเปรียบเทียบการทดลองทั้ง 2 วิธี คือ การปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อน และ การปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อนกับกลุ่มข้อมูลทดสอบ (ข) ที่แบ่งไว้ในช่วงแรกจำนวน 32,444 ชิ้น ได้ผลตามตารางที่ 4

รูปแบบการปรับพารามิเตอร์	จำนวนงานที่ดีในกลุ่มข้อมูลทดสอบ (ข)
ไม่มีกรปรับปรุงพารามิเตอร์	24607
ปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อน	27119
ปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อน	24746

จากตารางที่ 4 พบว่า การปรับพารามิเตอร์แบบที่ให้งานเสียมากที่สุดก่อนสามารถทำให้จำนวนงานดีเพิ่มขึ้น 10.2% และการปรับพารามิเตอร์แบบที่ให้งานเสียน้อยที่สุดก่อนสามารถทำให้จำนวนงานดีเพิ่มขึ้น 0.57%

5. สรุปผลการทดลอง

งานวิจัยนี้มุ่งพัฒนาคุณภาพของกระบวนการผลิตรถยนต์โดยการปรับปรุงค่าพารามิเตอร์ที่สามารถทำการปรับได้ ต้นไม้ตัดสินใจอัลกอริทึม C4.5 ถูกนำมาประยุกต์ใช้เพื่อหาค่าพารามิเตอร์ที่สำคัญในกระบวนการผลิต โดยต้นไม้ตัดสินใจ

เพื่อทำการจำแนกประเภทของผลิตภัณฑ์ออกเป็นงานดี และงานเสีย การทำการปรับปรุงค่าพารามิเตอร์เพื่อประมาณจำนวนของผลิตภัณฑ์ที่มีคุณภาพดีให้มีแนวโน้มไปในทางที่มากขึ้น ซึ่งงานวิจัยนี้แบ่งวิธีการปรับปรุงค่าพารามิเตอร์แบ่งออกเป็น 2 วิธี คือ การปรับจากพารามิเตอร์ที่ให้งานเสียมากที่สุดก่อน และ การปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อน โดยวิธีการปรับพารามิเตอร์แบบที่ให้งานเสียมากที่สุดก่อนเป็นวิธีที่ดีที่สุดเนื่องจากให้ค่าจำนวนงานดีเพิ่มขึ้นมากกว่าวิธีการปรับจากพารามิเตอร์ที่ให้งานเสียน้อยที่สุดก่อน จึงเหมาะสมที่จะนำไปสร้างแบบจำลองในการพัฒนาคุณภาพของฮาร์ดดิสก์

เอกสารอ้างอิง

- [1] The Board of Investment of Thailand (BOI), "Annual Report 2010," pp.52-53, 2010.
- [2] Mitchell, T.M., "Machine Learning", McGraw-hill, 1997.
- [3] Kijssirikul, B., "Data Mining Algorithms", the final report on the Joint Government and Private Sectors, Chulalongkorn University 2004 (in Thai).
- [4] Remco R. Bouckaert, Eibe Frank, Mark A. Hall, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. WEKA-experiences with a java open-source project. Journal of Machine Learning Research, 11:2533-2541, 2010.
- [5] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, pp.81-106, 1986
- [6] J. R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
- [7] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees. Wadsworth, 1984.
- [8] G. Kass, "An exploratory technique for investigating large quantities of categorical data. Applied Statistics," Appl.Statist., vol. 29, no. 2, pp. 119-127, 1980.
- [9] Taetragee U. and Achalakul T., "Applying Decision Tree in Fault Pattern Analysis for HGA Manufacturing", Int. Conf. on Complex, Intelligent and Software Intensive Systems, 2009, pp.83-89.
- [10] Yi-Ting Huang, Fan-Tien Cheng and Min-Hsiung Hung, "Developing a Product Quality Fault Detection Scheme", IEEE Int. Conf. on Robotics and Automation, Japan, May 12-17, 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, 2009, pp. 1211-1216.