การตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกในกระบวนการผลิตจานบันทึกแบบแข็ง

นายธนาธาร ภัทรกวิน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Detection of Machines Anomaly from Log Files in Hard Disk Manufacturing Process

Mr. Thanatarn Pattarakavin

5462082 18

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Software Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2016

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกในกระบวนการผลิตจานบันทึกแบบแข็ง
โดย	นายธนาธาร ภัทรกวิน
สาขาวิชา	วิศวกรรมซอฟต์แวร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

5462108218

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

.....กรรมการภายนอกมหาวิทยาลัย

(ดร.เฉลิมทรัพย์ สังขวิจิตร)

ธนาธาร ภัทรกวิน : การตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกในกระบวนการผลิตจานบันทึกแบบแข็ง (Detection of Machines Anomaly from Log Files in Hard Disk Manufacturing Process) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. ดร.ประภาส จงสถิตย์วัฒนา, 40 หน้า.

การผลิตฮาร์ดดิสก์เป็นอุตสาหกรรมที่สำคัญของประเทศไทย กระบวนการผลิตมีความซับซ้อนสูงและประกอบด้วยเครื่องจักรอัตโนมัติจำนวนมากทำงานการผลิตอย่างต่อเนื่อง เมื่อมีเหตุการณ์ผิดปกติเกิดขึ้นสายการผลิตจะหยุดการทำงาน และทีมวิศวกรจะวินิจฉัยโดยระบุและค้นหาแหล่งที่มาของข้อผิดพลาดในหมู่เครื่องจักรเหล่านั้น และแก้ไขให้ถูกต้องอย่างรวดเร็ว ในสายการผลิตเครื่องจักรอัตโนมัติทั้งหมดติดต่อกับเครื่องบริการโดยบันทึกเป็นแฟ้มไว้อย่างต่อเนื่อง วิศวกรใช้แฟ้มบันทึกเหล่านี้ในการวินิจฉัยสาเหตุของข้อผิดพลาด งานวิจัยนี้ใช้วิธีการเรียนรู้ของเครื่องจักรเพื่อแจ้งเหตุการณ์ที่เกิดความผิดปกติในแฟ้มล็อก ได้แก่ ความผิดปกติที่มาจากวิศวกร ความผิดปกติที่มาจากระบบ และความผิดปกติที่มาจากซอฟต์แวร์ ผลการทดลองแสดงให้เห็นว่าการตรวจสอบมีความถูกต้องมากถึงหนึ่งร้อยเปอร์เซ็นต์จากการวัดผลความแม่นยำด้วย 10-Fold validation และสามารถช่วยให้ทีมงานวิศวกรดำเนินการวินิจฉัยได้รวดเร็วและมีประสิทธิภาพ

5462082 18

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมซอฟต์แวร์ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2559

5770194621 : MAJOR SOFTWARE ENGINEERING

KEYWORDS: HARD DISK MANUFACTURING / HEAD STACK ASSEMBLY / LOG FILES / MACHINE LEARNING / DECISION TREE

THANATARN PATTARAKAVIN: Detection of Machines Anomaly from Log Files in Hard Disk Manufacturing Process. ADVISOR: PROF. PRABHAS CHONGSTITVATANA, Ph.D., 40 pp.

Hard disk manufacturing is an important industry in Thailand. The production line of its manufacturing process is highly complex and consists of hundreds of automated machines running a continuous flow production. When an anomaly event occurs, the production line has to be stopped and the diagnosis engineering team must identify and locate the source of error amongst those machines and correct them quickly. In an automated production line, all machines are monitored and their log files are sent to a server continuously. Engineers use these log files to diagnose the causes of the errors. This work proposes to use machine learning method to identify anomalous events in the log files such as anomaly caused by engineers, anomaly from systems, and anomaly from software. The experimental results showed that accuracy was 100% using 10-Fold validation, so it is very accurate and it can help teams of engineers to perform diagnosis quickly and effectively.

5462082I8

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

Field of Study: Software Engineering Advisor's Signature .....

Academic Year: 2016

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จเสร็จสมบูรณ์ได้ด้วยความกรุณาของท่านศาสตรจารย์ ดร. ประภาส จงสถิตวัฒนา ที่ได้สละเวลาให้คำแนะนำปรึกษาวิทยานิพนธ์อย่างสม่ำเสมอ ผู้วิจัยขอกราบขอบพระคุณอาจารย์เป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอขอบพระคุณผู้ช่วยศาสตรจารย์ ดร.สุกรี สินธุภิญโญ ประธานคณะกรรมการวิทยานิพนธ์ อาจารย์ผู้สอนวิชาการเรียนรู้ของเครื่องอีกทั้งเพิ่มเติมทักษะโปรแกรมภาษาไพทอน และขอขอบพระคุณ ดร.เฉลิมทรัพย์ สังขวิจิตร กรรมการจากภายนอก ที่เสียสละเวลาและกรุณาให้คำแนะนำวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณคณาจารย์สาขาวิชาวิศวกรรมซอฟต์แวร์และคณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ประสิทธิประสาทวิชาความรู้มาโดยตลอด

ขอขอบคุณ คุณเกียรติพงษ์ หล้านัก หัวหน้างานที่อนุญาตให้ดำเนินการวิจัย ดร. นเรนทร์ หล่อบรรจงสุข ผู้บริหารระดับสูงที่คอยให้การสนับสนุน และเพื่อนร่วมงานทุกคนที่ค่อยให้กำลังใจในการทำวิจัยด้วยดีตลอดมา

วิทยานิพนธ์ฉบับนี้ได้รับความอนุเคราะห์จาก Western Digital เพื่อสนับสนุนโครงการด้านการศึกษาและโครงการให้ความช่วยเหลือของมหาวิทยาลัย อย่างไรก็ตามวิทยานิพนธ์ฉบับนี้จัดทำขึ้นโดยผู้วิจัยเป็นผู้สร้างสรรค์ขึ้นเองและไม่ได้จัดพิมพ์เผยแพร่โดยหรือในนามของ Western Digital

สุดท้ายนี้ ผู้วิจัยขอขอบคุณครอบครัวที่เป็นกำลังใจและให้การสนับสนุนทำให้ผู้วิจัยสามารถทำวิจัยได้สำเร็จลุล่วง ตลอดจนเจ้าหน้าที่ของภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ให้การสนับสนุน ส่งเสริม และช่วยเหลือในการดำเนินการต่าง ๆ เสมอมา

5462082I8

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตของการวิจัย 3

1.4 ขั้นตอนการศึกษาและวิจัย 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 บทความจากวิทยานิพนธ์ 3

2.1 ทฤษฎีที่เกี่ยวข้อง 5

2.1.1 การผลิตฮาร์ดดิสก์ 5

ก. กระบวนการผลิตหัวอ่าน / เขียนข้อมูล (Slider Fabrication หรือ SF) 5

ข .กระบวนการประกอบหัวอ่าน / เขียนข้อมูล (Head Gimbal Assembly หรือ HGA) 5

ค. กระบวนการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ (Head Stack Assembly หรือ HSA) 6

ง. กระบวนการผลิตฮาร์ดดิสก์ไดร์ฟ (Hard Disk Drive หรือ HDD) 6

2.1.2 ชุดประกอบหัวอ่าน / เขียนสำเร็จ 7

2.1.3 แฟ้มล็อก 8

5462082 18

หน้า

2.1.4 การเรียนรู้ของเครื่อง (Machine Learning) 9
ก. การเรียนรู้แบบเบส์)Bayesian Learning) 9
ข. โครงข่ายประสาทเทียม (Neural Network) 9
ค .ต้นไม้ตัดสินใจ (Decision Tree) 10
ง. ตัวอย่างของวิธีการแบบอื่น ๆ 11
2.2 งานวิจัยที่เกี่ยวข้อง 11
บทที่ 3 แนวคิดและวิธีการดำเนินการวิจัย 14
3.1 เครื่องมือที่เกี่ยวข้อง 16
3.1.1 Orange Canvas 16
3.2 ข้อมูลเบื้องต้น (Preliminary Data) 17
บทที่ 4 การดำเนินงานวิจัย 21
4.1 การประมวลผลก่อน (Preprocessing) 21
4.2 การสร้างแบบจำลองและนำแบบจำลองที่ได้ไปใช้ (Model Construction and Model Usage) 24
4.3 การประเมินค่า (Evaluation) 24
4.4 การเรียนรู้ด้วยต้นไม้ตัดสินใจอย่างละเอียด 25
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 34
รายการอ้างอิง 35
ประวัติผู้เขียนวิทยานิพนธ์ 40

546208218

สารบัญตาราง

ตารางที่ 1 ผลการสำรวจความจุของแฟ้มล็อก เมื่อวันที่ 31 มกราคม ค.ศ.2015 17

ตารางที่ 2 คำอธิบายข้อมูลหลังการประมวลผลก่อน 23

ตารางที่ 3 เปรียบเทียบประสิทธิภาพของการเรียนรู้ของเครื่อง 4 วิธี 24

ตารางที่ 4 เปรียบเทียบประสิทธิภาพของการเรียนรู้ข้อมูลที่ผิดปกติ 4 วิธีการเรียนรู้ของเครื่อง 25

ตารางที่ 5 เปรียบเทียบประสิทธิภาพของการเรียนรู้ข้อมูลปกติ 4 วิธีการเรียนรู้ของเครื่อง 25

ตารางที่ 6 ประสิทธิภาพของการเรียนรู้ของเครื่องด้วยต้นไม้ตัดสินใจ 29

ตารางที่ 7 ประสิทธิภาพของการเรียนรู้ของเครื่องกับข้อมูลที่ผิดปกติด้วยต้นไม้ตัดสินใจ 30

ตารางที่ 8 ประสิทธิภาพของการเรียนรู้ของเครื่องกับข้อมูลปกติด้วยต้นไม้ตัดสินใจ 30

5492208218

สารบัญภาพ

ภาพที่ 1 Hype Cycle for Emerging Technologies, 2015 2

ภาพที่ 2 แบบจำลองกระบวนการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ 6

ภาพที่ 3 ชุดประกอบหัวอ่าน / เขียนสำเร็จ 7

ภาพที่ 4 ซอฟต์แวร์สนับสนุนการผลิตที่บริการเครื่องจักรต่าง ๆ 8

ภาพที่ 5 การจำลองตัวอย่างของแฟ้มล็อก 8

ภาพที่ 6 ตัวอย่างของการใช้โครงข่ายประสาทเทียม 10

ภาพที่ 7 ตัวอย่างการใช้ต้นไม้ตัดสินใจ 10

ภาพที่ 8 กระบวนการประกอบหัวอ่าน/เขียนสำเร็จ 12

ภาพที่ 9 รูปแบบการทำงานในปัจจุบัน 14

ภาพที่ 10 รูปแบบการทำงานเพิ่มเติมจากการอ่านแฟ้มล็อก 15

ภาพที่ 11 การสำรวจเครื่องบริการที่เลือกมาใช้งานวิจัยนี้ 16

ภาพที่ 12 ตัวอย่างการนำชุดข้อมูลของสวนสัตว์เข้าโปรแกรม Orange Canvas 17

ภาพที่ 13 ตัวอย่างการแยกกลุ่มข้อมูลและสร้างมโนภาพด้วยโปรแกรม Orange Canvas 17

ภาพที่ 14 การสำรวจเครื่องบริการที่บันทึกแฟ้มล็อก 18

ภาพที่ 15 กิจกรรมแนวคิดในการดำเนินงานวิจัย 20

ภาพที่ 16 ตัวอย่างข้อมูลที่ถูกประมวลผลก่อน 23

ภาพที่ 17 การทดลองเพื่อแยกกลุ่มข้อมูลปกติและผิดปกติโดย Orange Canvas 29

ภาพที่ 18 ต้นไม้ตัดสินใจเพื่อแยกกลุ่มข้อมูลปกติและผิดปกติ 31

ภาพที่ 19 ต้นไม้สินใจที่ตัดคุณลักษณะการนับคำว่า “No Data” ออก 32

ภาพที่ 20 ต้นไม้สินใจที่ตัดหมายเลขข้อความออก 33

ภาพที่ 21 ผลลัพธ์ของการนำแบบจำลองมาใช้จำแนกแฟ้มล็อก 33

5462082218

บทที่ 1
บทนำ

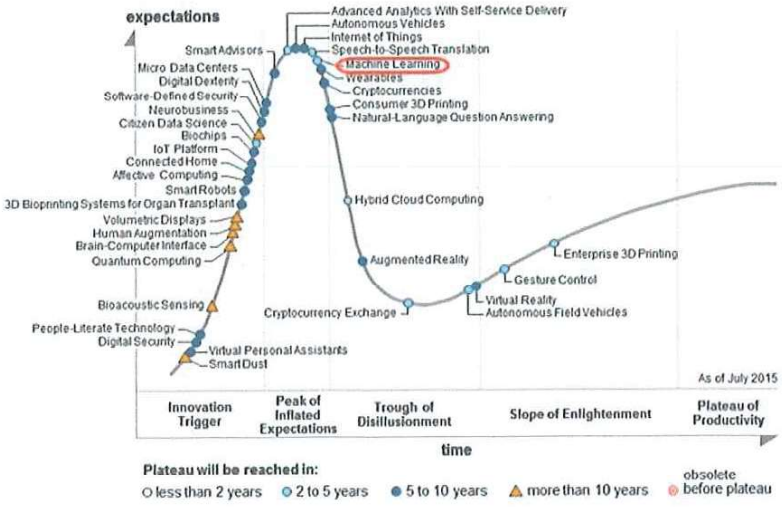
1.1 ความเป็นมาและความสำคัญของปัญหา

การผลิตจานบันทึกแบบแข็งเป็นหนึ่งในกลุ่มอุตสาหกรรมผลิตเพื่อการส่งออกของประเทศไทย [1, 2] สร้างรายได้ถึง 15,000 ล้านดอลลาร์สหรัฐของมูลค่าการส่งออกและทำให้เกิดการจ้างงานกว่า 110,000 คนตั้งแต่ปี ค.ศ.2007 [3] บริษัท เวสเทิร์นดิจิตอล เป็นหนึ่งในบริษัทที่ผลิตจานบันทึกแบบแข็ง โดยติดห้าร้อยอันดับแรกของนิตยสารฟอร์จูนในปี ค.ศ.1997 [4]และอยู่ในอันดับที่ดีขึ้นในห้าร้อยอันดับแรกของนิตยสารดังกล่าวในปี ค.ศ.2015 [5] โดยมีประเทศไทยเป็นฐานการผลิตที่ใหญ่ที่สุดในโลก [6] และบริษัท เวสเทิร์นดิจิตอล จำกัด (ประเทศไทย) เป็นผู้ปฏิบัติที่เป็นเลิศ (Best Practice)[7] ในสมาชิกของโครงการความร่วมมือระหว่างคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยกับภาคอุตสาหกรรม (Industrial Liaison Program หรือ ILP) [8]

5462082218

การประกอบชุดอ่าน เขียนสำเร็จ / (Head Stack Assembly หรือ HSA) อยู่ในกระบวนการผลิตในโรงงานอันเต็มไปด้วยเครื่องจักรนับพันทำงานกันอย่างหลากหลายสลับซับซ้อนเต็มไปด้วยซอฟต์แวร์มากมาย โดยซอฟต์แวร์ที่คอยให้บริการซอฟต์แวร์ตัวอื่น ๆ นั้นจะทำงานบนเครื่องบริการตลอดเวลาและมีการปรับปรุงอย่างต่อเนื่อง เพื่อตอบสนองความต้องการของตลาด โดยเขียนลงบันทึก (Log) ดุจดังรอยเท้า (Footprint) ของซอฟต์แวร์ที่สามารถย้อนรอยและนำมาตรวจสอบโดยปัจจุบันวิศวกรจะเข้ามาตรวจสอบแฟ้มล็อกเมื่อเกิดปัญหาขึ้นแล้ว

แฟ้มล็อกจากเครื่องจักรต่าง ๆ มีรูปแบบที่แตกต่างกัน [9] และมีข้อความที่ไม่สำคัญจำนวนมากที่รบกวนการวิเคราะห์ ดังนั้น การวิเคราะห์ปัญหาจากเครื่องจักรที่มีความซับซ้อนจำเป็นต้องอาศัยผู้เชี่ยวชาญ ข้อผิดพลาดมากมายเป็นผลจากพนักงานคุมเครื่องจักร ยกตัวอย่างพนักงานคุมเครื่องที่ละเว้นข้อความที่แสดงผลบนหน้าจอสำหรับเครื่องฉายรังสีเอกซ์เครื่องใหม่ เนื่องจากมีข้อความเตือนมากมายนับสิบนับร้อยหน้ากระดาษเอสี่และบางข้อความเป็นเรื่องปกติ แต่ในบางข้อความที่ละเลยไปนั้นเกี่ยวข้องกับการตั้งค่าพลังงานรังสีเอกซ์ ส่งผลให้เครื่องส่งพลังงานเต็มกำลัง แทนที่การบำบัดกลับกลายเป็นทำให้ผู้รับการบำบัดเสียชีวิต 4 ราย ซึ่งมาพบภายหลังว่าสาเหตุเกิดมาจากการละเลยข้อความที่สำคัญเหล่านั้น [10]



ภาพที่ 1 Hype Cycle for Emerging Technologies, 2015

5462092218

บริษัทการ์ตเนอร์นำเสนอพัฒนาการของเทคโนโลยีในปี ค2015.ศ. ผ่านกราฟเรียกว่า Hype Cycle ในภาพที่ 1 โดยจัดการเรียนรู้ของเครื่องอยู่ในระยะที่ 2 (Peak of Inflated Expectations) เป็นระยะที่เทคโนโลยีดังกล่าวเป็นที่รับรู้มากขึ้น โดยเป็นพื้นฐานให้กับ 10 แนวโน้มเทคโนโลยีสำหรับปี ค.ศ.2016 [11] ในลำดับที่ 5 คือการเรียนรู้ของเครื่องขั้นสูง (Advanced Machine Learning) และลำดับที่ 6 คือ ผู้ช่วยและสิ่งของที่สามารถเรียนรู้และโต้ตอบได้อย่างอัตโนมัติ (Autonomous Agents and Things) ซึ่งเทคโนโลยีเหล่านี้จะเป็นเทคโนโลยีที่ถูกพัฒนาต่อเนื่องยาวนานไปอีกถึง 20 ปี

การเรียนรู้ของเครื่อง (Machine Learning) ซึ่งเป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) โดยมุ่งเน้นที่จะพยายามสอนคอมพิวเตอร์ให้เรียนรู้ได้เหมือนมนุษย์ วิธีการเรียนรู้ของเครื่องมีประสิทธิผลในการบ่งบอกคุณสมบัติและแพทเทิร์นที่เป็นประโยชน์จากชุดข้อมูล (data set) [12] หากนำการเรียนรู้ของเครื่องเข้ามาทำหน้าที่ทดแทนมนุษย์ในการอ่านและวิเคราะห์แฟ้มล็อกดังกล่าวอย่างอัตโนมัติก็จะช่วยลดค่าใช้จ่ายให้แก่บริษัทได้

1.2 วัตถุประสงค์ของการวิจัย

วัตถุประสงค์งานวิจัยชิ้นนี้คือสร้างเครื่องมือตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกในกระบวนการผลิตฮาร์ดดิสก์ในส่วนของการประกอบชุดอ่าน / เขียนสำเร็จ (HSA)

1.3 ขอบเขตของการวิจัย

พัฒนาเครื่องมือวิเคราะห์แฟ้มล็อกว่าเครื่องจักรทำงานปกติหรือไม่ ผ่านการเรียนรู้ของเครื่องโดยการศึกษาแฟ้มล็อกของซอฟต์แวร์บนเครื่องบริการ โดยครอบคลุมงานบำรุงรักษาซอฟต์แวร์ที่สนับสนุนกระบวนการผลิตฮาร์ดดิสก์ในส่วนของการประกอบหัวอ่าน/เขียน (HSA)

1.4 ขั้นตอนการศึกษาและวิจัย

1.4.1 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1.4.2 ศึกษาขั้นตอนการทำงานของแฟ้มล็อก

1.4.3 ออกแบบวิธีการการเรียนรู้ของเครื่องที่ใช้งาน

1.4.4 ทำการทดลองตามวิธีที่นำเสนอ

1.4.5 วิเคราะห์ผลการทดลอง

1.4.6 สรุปผลการวิจัย

5462092118

1.5 ประโยชน์ที่คาดว่าจะได้รับ

เครื่องมือที่วิเคราะห์แฟ้มล็อกของซอฟต์แวร์ที่สนับสนุนการผลิตชุดประกอบหัวอ่าน / เขียนสำเร็จอย่างอัตโนมัติ และสามารถพัฒนาเป็นต้นแบบในการสร้างเครื่องมือช่วยทดแทนงานที่ปฏิบัติด้วยวิศวกรได้ในอนาคต

1.6 บทความจากวิทยานิพนธ์

1.6.1 บทความชื่อ "Detection of Machines Anomaly from Log Files in Hard Disk Manufacturing Process" ได้รับการตอบรับให้ตีพิมพ์และนำเสนอแล้วต่อที่ประชุมวิชาการ 2016 7th International Conference on Manufacturing Science and Technology (ICMST 2016) ระหว่างวันที่ 16 – 18 กรกฎาคม ค.ศ. 2016 ที่มหาวิทยาลัยเคอทิน (Curtin University) มิริ (Miri) ซาราวัค (Sarawak) ประเทศมาเลเซีย (Malaysia) โดยมีการปรับปรุงและได้รับการตอบรับตีพิมพ์ต่อที่ประชุมวิชาการ 2016 International Conference on Multimedia Systems and Signal Processing (ICMSSP 2016) ระหว่างวันที่ 3 – 5 กันยายน ค.ศ. 2016 จัดโดยมหาวิทยาลัยตัมกั่ง (Tamkang University) นิวไทเป (New Taipei) ณ สาธารณรัฐจีนไต้หวัน (Taiwan)

1.6.2 บทความชื่อ "การตรจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกในกระบวนการผลิตฮาร์ดดิสก์ด้วยต้นไม้ตัดสินใจ (Detection of Machines Anomaly from Log File in Hard Disk Manufacturing Process using Decision Tree)" ได้รับตอบรับเพื่อตีพิมพ์ในในวารสารไทยการวิจัยดำเนินงาน (Thai Journal of Operations Research) ปีที่ 4 ฉบับที่ 2 (กรกฎาคม –

ธันวาคม พ.ศ. 2559) โดยวารสารไทยการวิจัยดำเนินงาน (Thai Journal of Operations Research) ถูกนำเข้าสู่ฐานข้อมูลศูนย์ดัชนีการอ้างอิงวารสารไทย (Thai Journal Citation Index-หรือ TCI) และได้รับการจัดกลุ่มวารสารที่ผ่านการรับรองคุณภาพของ TCI อยู่ในวารสารกลุ่มที่ 2

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การผลิตฮาร์ดดิสก์

โรงงานผลิตจานบันทึกแบบแข็งหรือฮาร์ดดิสก์จัดอยู่ในกลุ่มของอุตสาหกรรมการผลิต โดยจะมีความมุ่งมันในการผลิตผลิตภัณฑ์ให้มีคุณภาพสูงในต้นทุนการผลิตที่ต่ำ ผ่านแนวคิดแนวปฏิบัติของ 5 ส. และระบบการผลิตแบบลีน เป็นต้น เพื่อทำให้คุณภาพสูงขึ้นเรื่อย ๆ จนอยู่ในระดับที่ลูกค้าพึ่งพอใจ และพัฒนาให้ต้นทุนการผลิตต่ำลงเพื่อให้มีกำไร อันสร้างความเชื่อมั่นแก่ผู้ลงทุนมากสุด

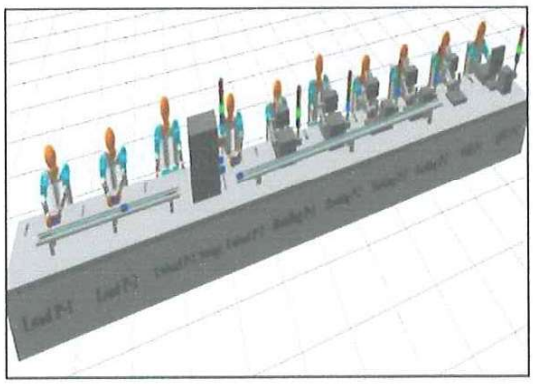
ระบบสายการผลิต ประกอบไปด้วย 4 ส่วน [6] คือ

ก. กระบวนการผลิตหัวอ่าน / เขียนข้อมูล (Slider Fabrication หรือ SF) คือขั้นส่วนสำคัญขั้นแรกในกระบวนการผลิตฮาร์ดดิสก์ ซึ่งกระบวนการผลิตเริ่มจากการตัดเวเฟอร์จากแผ่นใหญ่ให้อยู่ในรูปบาขนาดยาวประมาณ 2 นิ้ว แล้วนำไปขัดหน้างานเพื่อให้ค่าทางไฟฟ้าและพื้นผิวตามต้องการ หลังจากนั้นชิ้นงานจะถูกส่งไปยังกระบวนการขึ้นรูปผิวหน้าหัวอ่านหรือที่เรียกว่าสไลด์เดอร์ (Slider) ซึ่งประกอบด้วยการเคลือบพื้นผิวหน้าสไลด์เดอร์ด้วยซิลิคอลและคาร์บอนต่อด้วยขบวนการการถ่ายภาพ (Photography) และกลวิธีพิมพ์กัดกรด (Etching) ซึ่งการสร้างรูปแบบของหน้าแอร์แบริ่ง เซอเฟส (Air Bearing Surface) เพื่อให้ระดับการบินของหัวอ่านการบินของฮาร์ดดิสก์อยู่ ณ ระดับที่ถูกต้อง จากนั้นชิ้นงานจะถูกตรวจสอบค่าทางไฟฟ้าและต่อด้วยขบวนการตัดแปะให้อยู่ในขนาดที่เท่ากัน หัวทุกตัวจะถูกตรวจสอบกระบวนการขั้นสุดท้ายและทำความสะอาดก่อนส่งต่อไปยังหน่วยงานผลิตหัวอ่าน / เขียนข้อมูลต่อไป

ข .กระบวนการประกอบหัวอ่าน / เขียนข้อมูล (Head Gimbal Assembly หรือ HGA) [13] เริ่มกระบวนการผลิตจากการถ่ายโอนตัวซัสเทนชันแอสเซมบลีสะอาดด้วยเครื่อง Acquire cleaning และ ตรวจสอบความสมบูรณ์ของคุณภาพขั้นสุดท้าย ก่อนส่งไปประกอบเป็นชุดหัวเขียน / สำเร็จต่อไป

9473028181

ค. กระบวนการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ (Head Stack Assembly หรือ HSA) [13] เป็นอุปกรณ์ที่บันทึกและอ่านข้อมูลจากจานแม่เหล็กในฮาร์ดดิสก์ไดร์ฟ วิธีการประกอบชุดหัวอ่าน / เขียนสำเร็จ (Head Stack) คือการนำหัวอ่าน / เขียนข้อมูลมาประกอบให้เข้ากับแอคโชเอเตอร์ พลีเวิร์ด เฟล็ค แอสเซมบลี (Actuator Pivot Flex Assembly หรือ APFA) โดยใช้เครื่องมือที่มีประสิทธิภาพสูง เช่น เครื่องประกอบแบบบีดอัด)Swage Press Machine) ไฮบอนบอนดิ่ง แมชชีน (High bond bonding Machine) แล้วทำการตรวจสอบคุณสมบัติทางกายภาพด้วยกล้องขยาย ต่อจากนั้นจะทำการตรวจสอบคุณสมบัติทางไฟฟ้าด้วยเครื่องมือที่เรียกว่า ควอไซด์สเตติก เทสเตอร์ (Quasi Static Tester) ทำการทดสอบการเขียนและอ่านของชุดประกอบหัวอ่าน / เขียนสำเร็จเพื่อประกอบเป็นฮาร์ดดิสก์ไดร์ฟต่อไปอย่างละเอียดก่อนส่งมอบดังภาพที่ 2

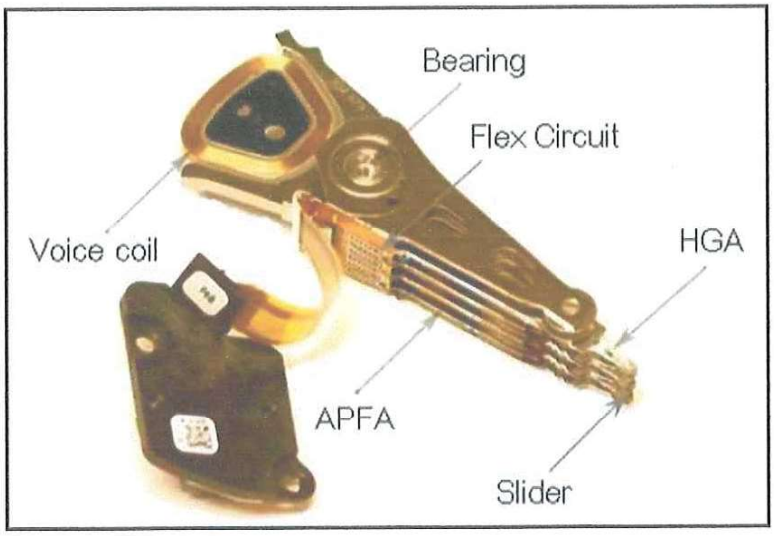


ภาพที่ 2 แบบจำลองกระบวนการประกอบชุดประกอบหัวอ่าน/เขียนสำเร็จ [3]

ง. กระบวนการผลิตฮาร์ดดิสก์ไดร์ฟ (Hard Disk Drive หรือ HDD) เริ่มกระบวนการผลิตจากการอ่านข้อมูลของพนักงาน เครื่องจักร และชิ้นส่วนที่จะทำการประกอบลงในระบบ MITECS (Manufacturing Information Technology Execution Control System) และจะเริ่มกระบวนการผลิตจากการเขียนสัญญาณ Servo ลงบนแผ่น Media โดยเครื่อง Multi Disk Writer (MDW) จากนั้นจึงทำการประกอบแผ่นมีเดีย (Media) แรมป์(Ramp) HSA Voice Coil Motor (VCM) แล้วจึงประกอบ Top Cover เข้ากับหัวอ่านเขียนข้อมูลก่อนส่งไปยังแบคเอน คอนเวเยอร์ (Backend Conveyer) เพื่อประกอบ Printed Circuit Boards Assembly (PCBAX) เข้ากับหัวอ่าน / เขียนข้อมูล แล้วทดสอบฟังก์ชันการทำงานของฮาร์ดดิสก์ไดร์ฟด้วยเครื่องเอ็กคาลิเบอร์ (Xcalibre) จนได้ฮาร์ดดิสก์ไดฟ์ที่มีคุณภาพก่อนทำการบรรจุผลิตภัณฑ์ เพื่อทำการจัดส่งให้แก่ลูกค้าต่อไป

5462082218

2.1.2 ชุดประกอบหัวอ่าน / เขียนสำเร็จ

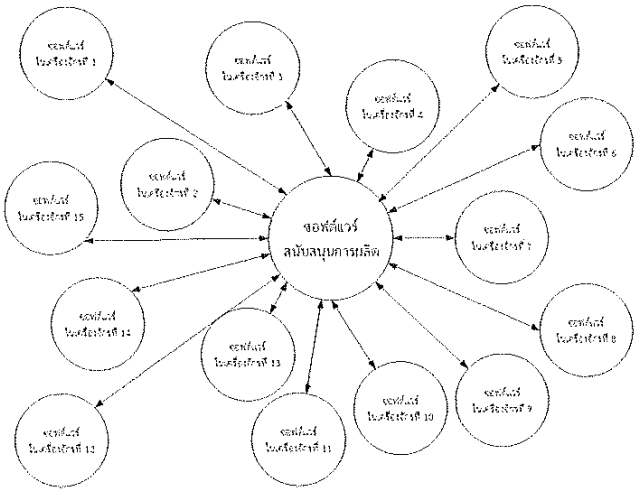


ภาพที่ 3 ชุดประกอบหัวอ่าน / เขียนสำเร็จ [3]

5476202818

จากภาพที่ 3 แสดงให้เห็นภาพชุดประกอบหัวอ่าน / เขียนสำเร็จอันมีชิ้นส่วนอื่นมาประกอบกันขึ้น อาทิ เช่น สไลด์เดอร์ เอชจีเอและเอพีเอฟเอ โดยผ่านกระบวนการผลิตผ่านเครื่องจักรมากมายหลายหลากจนเป็นชิ้นงานดังภาพก่อนจะนำส่งไปเป็นชิ้นส่วนหนึ่งในการประกอบเป็นฮาร์ดดิสก์ต่อไป

การสนับสนุนการผลิตถือว่าเป็นหัวใจของโรงงานผลิตฮาร์ดดิสก์ แล้วในการผลิตก็มีซอฟต์แวร์ที่สนับสนุนการผลิตที่ทำงานตลอดเวลาอยู่บนเครื่องบริการที่ให้บริการโดยจะมีการเขียนลงบันทึก (Log) เป็นแฟ้มข้อมูล (File) เรียกว่าแฟ้มล็อก (Log Files) ทุก ๆ รายการเปลี่ยนแปลง (Transaction) ตลอดเวลาเป็นกระแสข้อมูล (Stream) อันสะท้อนถึงสุขภาพของซอฟต์แวร์ที่สนับสนุนการผลิต เมื่อจำเพาะลงบนงานบำรุงรักษาซอฟต์แวร์ที่สนับสนุนกระบวนการผลิตฮาร์ดดิสก์ในส่วนของการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ สามารถอธิบายแทนด้วยวงกลมตรงกลางของภาพที่ 4



ภาพที่ 4 ซอฟต์แวร์สนับสนุนการผลิตที่บริการเครื่องจักรต่าง ๆ

2.1.3 แฟ้มล็อก

5462082218

แฟ้มล็อกแฟ้มนามสกุล “log” ขนาดตั้งแต่ 1 กิโลไบต์ไปจนถึง 5 เมกะไบต์ต่อแฟ้ม ถูกบันทึกจากซอฟต์แวร์ที่สนับสนุนการผลิตการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ โดยจำลองดังภาพที่ 5

logfile.log - Notepad

File Edit Format View Help

29-12-2015 09:10:07.438: R: 1,17,ABCDEFGHIJKLM,0,12345B,AB.01.234,A,BC.D1.2.E.FG.HI.JK

,L,M,,,

29-12-2015 09:10:07.438: A: 1,17,ABCDEFGHIJKLM,0,12345B,AB.01.234,A,BC.D1.2.E.FG.HI.JK,

L,M,,,

29-12-2015 09:10:07.453: A: CheckMachineInitialHSA - Start

29-12-2015 09:10:07.782: A: CheckMachineInitialHSA - End

29-12-2015 09:10:07.813: S:

1,0,ABCDEFGHIJKLM,0,1,A,0,BC.D1.2.E.FG.HI.JK,AB.01.234,GREEN0A,9L,ABCABCABCABCABCABCABC1230

ABCDEF,AB0123456789AB1CD4EFG43H0IJKLMN,,0,A12B345678901A0123456B,1,14,,,,,2;3,,,,,,,,,,,,,,,,,,,,,,

,,,,,,,,,,,,,,,,,,,,

29-12-2015 09:10:07.813: 1, Response Time: 0.1234567 seconds.

ภาพที่ 5 การจำลองตัวอย่างของแฟ้มล็อก

2.1.4 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องเป็นการตั้งคำถามเพื่อหาวิธีที่สามารถสร้างระบบคอมพิวเตอร์ที่ปรับปรุงโดยอัตโนมัติด้วยประสบการณ์และกฎพื้นฐานที่ควบคุมกระบวนการเรียนรู้ทั้งหมด [8] ตัวอย่างวิธีการเรียนรู้ของเครื่อง 4 กลุ่ม คือ

ก. การเรียนรู้แบบเบส์ (Bayesian Learning)

วิธีการเรียนรู้แบบเบส์ใช้หลักการของความน่าจะเป็นโดยมีสมมติฐานว่าการประมาณของความสนใจขึ้นอยู่กับการกระจายความน่าจะเป็น (probability distribution) [12] อันเป็นปฐมฤกษ์ของการเรียนรู้ของเครื่องและใช้งานจริงได้ดี ยกตัวอย่างเช่น นาอีฟเบส์หรือการเรียนรู้เบส์อย่างง่าย (Naïve Bayes Learning) [14] เป็นวิธีการจำแนกประเภทข้อมูลที่ง่าย เร็ว

ทฤษฎีของเบส์ (Bayes theorem)

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (1)$$

โดยที่

$P(h)$ คือ ความน่าจะเป็นก่อนหน้าของสมมติฐาน h

$P(D)$ คือ ความน่าจะเป็นก่อนหน้าของชุดข้อมูลตัวอย่าง D

$P(h|D)$ คือ ความน่าจะเป็นของสมมติฐาน h เมื่อรู้ชุดข้อมูลตัวอย่าง D

$P(D|h)$ คือ ความน่าจะเป็นของชุดข้อมูลตัวอย่าง D เมื่อรู้สมมติฐาน h

$$h_{MAP} = \underset{h \in H}{argmax}\ P(h|D) \quad (2)$$

$$= \underset{h \in H}{argmax}\ \frac{P(D|h)P(h)}{P(D)} \quad (3)$$

$$= \underset{h \in H}{argmax}\ P(D|h)P(h) \quad (4)$$

$$h_{ML} \equiv \underset{h \in H}{argmax}\ P(D|h) \quad (5)$$

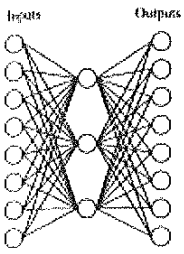
สมการที่ 1 สามารถแทนด้วยสมการที่ 2 สมการที่ 3 และสมการที่ 4 จนเป็นสมการที่ 5

ข. โครงข่ายประสาทเทียม (Neural Network)

Neural Network หรือ Artificial Neural Network อธิบายภาพโหนดแต่ละตัวคือโหนดของการคำนวณ วิถีแต่ละเส้นมีน้ำหนัก ซึ่งสามารถคำนวณประกอบเป็นฟังก์ชันได้ แบ่งออกเป็น 3 ชั้นใหญ่ คือ ชั้นขาเข้า (Input) ชั้นที่ซ้อนเร้น (Hidden) และชั้นขาออก (Output) โดยมี

546203218

ข้อจำกัดคือฝึกสอนช้า แต่มีจุดเด่นที่ใช้งานได้เร็ว โดยมีตัวอย่างการประยุกต์ใช้กับการเข้ารหัสเลขฐานสองด้วยตัวเลขแปดบิตดังภาพที่ 6



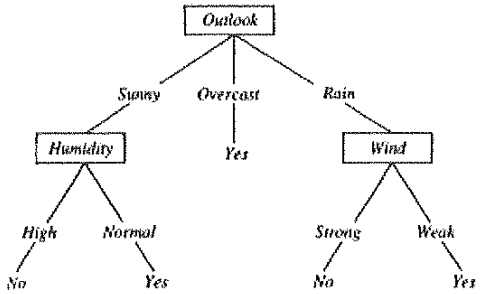
Input		Hidden Values				Output
10000000	→	.89	.04	.08	→	10000000
01000000	→	.15	.99	.99	→	01000000
00100000	→	.01	.97	.27	→	00100000
00010000	→	.99	.97	.71	→	00010000
00001000	→	.03	.05	.02	→	00001000
00000100	→	.01	.11	.88	→	00000100
00000010	→	.80	.01	.98	→	00000010
00000001	→	.60	.94	.01	→	00000001

ภาพที่ 6 ตัวอย่างของการใช้โครงข่ายประสาทเทียม [14]

ค .ต้นไม้ตัดสินใจ (Decision Tree)

เป็นการถามเพื่อแยกกลุ่มข้อมูล เพื่อลดเอนโทรปี (Entropy) คือความสับสนหรือตัวไม่

546208218



ภาพที่ 7 ตัวอย่างการใช้ต้นไม้ตัดสินใจ [14]

ภาพที่ 7 เป็นภาพตัวอย่างของต้นไม้ตัดสินใจ จากกรณีเล่นเทนนิสโดยแยกกลุ่มว่าใช่หรือไม่ใช่ เพื่อตอบคำถามว่าเช้าวันเสาร์สภาพอากาศเหมาะสมต่อการเล่นเทนนิสหรือไม่

โดยมีการแบ่งออกเป็น 3 ส่วนคือ

1) Decision nodes คือ โหนดแรกที่แยกกลุ่มข้อมูลเพื่อลดความสับสนลงให้ได้มากสุด

2) Chance nodes คือ โหนดถัด ๆ ลงมาจากโหนดแรก

3) End nodes คือ โหนดสุดท้ายของแต่ละกิ่ง

โดยมีอัลกอริทึมที่นิยมกันอยู่ 6 ตัว คือ

1) Iterative Dichotomiser 3 (ID3)

2) C 4.5 ปรับปรุงจาก ID3 ในการลดความความลำเอียงของข้อมูล

3) CART ย่อมาจาก Classification and Regression Tree

4) CHAID (CHI – squared Automatic Interaction Detector)

5) MARS ขยายความสามารถของ Decision trees ให้ทำงานกับข้อมูลที่เป็นตัวเลขได้ดียิ่งขึ้น

6) Conditional Inference Trees ใช้สถิติเป็นฐาน

ง. ตัวอย่างของวิธีการแบบอื่น ๆ

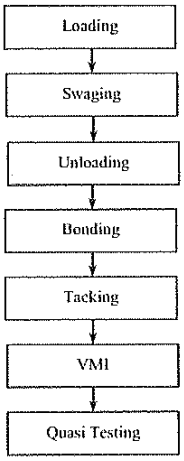
ตัวอย่างของวิธีการแบบอื่น ๆ เช่น เคเอ็นเอ็น (k-Nearest Neighbor หรือ k-NN) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine หรือ SVM) เป็นต้น เคเอ็นเอ็นเป็นการหาเพื่อนบ้านใกล้สุดว่ามีความคล้ายกับโหนดใกล้เคียงจำนวน k ตัว และสามารถประยุกต์ใช้หาข้อมูลที่ขาดหายหรือไม่รู้ค่าได้ ส่วนซัพพอร์ตเวกเตอร์แมชชีน เป็นเทคนิคในการทำจัดกลุ่ม (Classification) และทำนายผลได้ดี กล่าวคือแบ่งกลุ่มออกจากกันด้วยไฮเพอร์เพลน โดยแยกแยะกลุ่มข้อมูลได้ดีที่สุด

5492309218

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 บทความของวีรพล แก้วกล้าและเกียรติฟ้า ตั้งใจจิต (2010) ได้มีการกล่าวถึงคำศัพท์ที่เกี่ยวของกับกระบวนการผลิตฮาร์ดดิกส์ทั้งภาษาไทยและภาษาอังกฤษ [13, 15] อาทิเช่น HGA ย่อมาจาก Head Gimbal Assembly แทนภาษาไทยด้วยคำว่าการประกอบหัวอ่าน / เขียนข้อมูล และ HSA ย่อมาจาก Head Stack Assembly แทนภาษาไทยด้วยคำว่าชุดประกอบหัวอ่าน / เขียนสำเร็จ เป็นต้น

2.2.2 บทความของอมรเทพ ดอกไม้และอรรถกร เก่งพล (2010) แสดงให้เห็นภาพของชุดประกอบหัวอ่าน / เขียนสำเร็จว่าเป็นอุปกรณ์มีลักษณะอย่างไร ประกอบไปด้วยส่วนประกอบใดบ้าง [3] ตลอดจนกระบวนการประกอบต้องผ่านกระบวนการดังภาพที่ 8



ภาพที่ 8 กระบวนการประกอบหัวอ่าน/เขียนสำเร็จ [3]

2.2.3 บทความของ Stephen G. Eick, Michael C. Nelson และ Jerry D. Schmidt (1994) แสดงถึงการแปลงแฟ้มล็อกเป็นแผนภาพเพื่อให้ง่ายแก่การวิเคราะห์ผล [9] และเมื่อมาพิจารณาถึงการเรียนรู้ของเครื่อง อโณทัย ศิลเทพาเวทย์ สุกรี สินธุภิญโญ และประภาส จงสถิตย์วัฒนา (2012) ได้นำเสนอการนำการเรียนรู้ของเครื่องเข้ามาพัฒนาคุณภาพของการผลิตฮาร์ดดิสก์ โดยใช้วิธีการเรียนรู้ของเครื่องด้วยต้นไม้ตัดสินใจ [16], [17]

2.2.เครื่องมือที่เกี่ยวข้อง เช่น Weka [18] ย่อมาจากคำว่า Waikato Environment for Knowledge Analysis เป็นชุดของเครื่องมือสำหรับเรียนรู้ของเครื่องที่ได้รับการพัฒนาบนพื้นฐานของภาษาจาวา (Java) สามารถประมวลผลและทำงานได้หลายระบบปฏิบัติการ โดยสามารถรองรับข้อมูลได้หลายรูปแบบ เช่น CSV และแฟ้มข้อความ เป็นต้น โดยมีจุดเด่นด้านการใช้ทรัพยากรน้อยและไม่เสียค่าใช้จ่ายเนื่องจากเป็นโอเพนซอร์ซ ถูกนำประกอบบทเรียนและตำรา [19] โดยอาจารย์ Ian (2013) และอีกเครื่องมือหนึ่งคือ Orange Canvas เป็นโอเพนซอร์ซประกอบด้วยชุดเครื่องมือสำหรับเรียนรู้ของเครื่องพร้อมด้วยชุดการทำเหมืองข้อมูลสำหรับการวิเคราะห์ข้อมูล (Data Analysis) ผ่านโปรแกรมสคริปภาษาไพทอน (Python) ในรูปแบบโปรแกรมการแสดงผล ซึ่งถูกนำมาใช้ในบทความของ Miha Stajdohar และ Janez Demsar (2013) เพื่อวิเคราะห์ข้อมูลเครือข่ายผ่านการเรียนรู้ของเครื่องหลายวิธี [20] เช่น ต้นไม้ตัดสินใจและเคเอ็นเอ็น เป็นต้น

2.2.5 Tingting Gao, Chunling Du, Weijie Sun และ Lihua Xie (2016) นำเสนอว่ากระบวนการผลิตเต็มไปด้วยสิ่งรบกวน แม้แต่ผลลัพธ์ของกระบวนการผลิตก็อาจมารบกวนกระบวนการขาเข้าได้เช่นกัน [21] มีงานวิจัยเสนอการตรวจสอบเหตุการณ์ที่เกิดความผิดพลาดในแฟ้มล็อกโดยอัตโนมัติคือเป็นระบบที่สามารถตรวจสอบแฟ้มล็อกได้อย่างทันทีทันใด และแจ้งทีมวิศวกรเมื่อมีความผิดปกติเกิดขึ้น อีกผลงานหนึ่งของ Franka Schuster, Andreas Paul และ

Hartmut König (2013) ประยุกต์ใช้การตรวจสอบความผิดปกติในเครือข่ายระบบควบคุม [22] สามารถระบุตำแหน่งและต้นเหตุที่ก่อให้เกิดความผิดปกติ เครื่องมือนี้สามารถช่วยสนับสนุนวิศวกรในการหาสาเหตุที่ส่งผลกระทบต่อสายการผลิตและแก้ไขได้อย่างรวดเร็ว นอกจากนี้เทคนิคการเรียนรู้ของเครื่องยังสามารถเพิ่มพูนประสิทธิภาพของระบบ โดยอาศัยประสบการณ์จริง [14, 23] และยังมีการใช้การวิเคราะห์โดยใช้ต้นไม้ตัดสินใจเพื่อปรับปรุงประสิทธิภาพการผลิต [17] ดังนั้น การเรียนรู้ของเครื่องสามารถแยกความผิดปกติจากสิ่งรบกวน และช่วยลดกำลังคนในงานที่เกี่ยวข้องได้

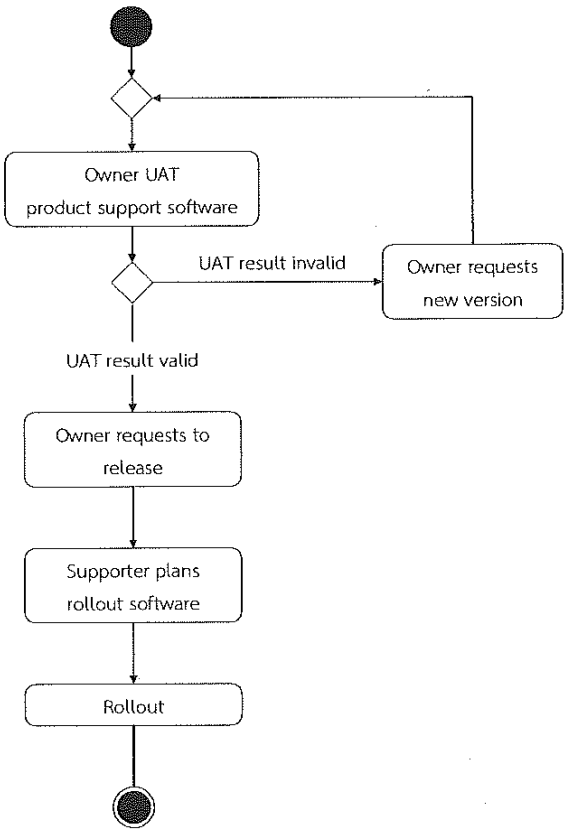
2.2. บทความของ 6Bernhard Schöölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, Robert C. Williamson (2001) ใช้ Supervised Learning สำหรับการรู้จำลายมือว่าเป็นตัวเลขใด ระหว่าง 0 ถึง 9[24] โดยปรับปรุงผลงานเดิมจากปี ค 2000 .ศ.ในการบอกความผิดปกติหรือ abnormal class และบทความของ Volker Roth (2004) และ Volker Roth (2006) ใช้ Unsupervised Learning ในการรู้จำใบหน้าระหว่างคนใส่แว่นธรรมดากับคนใส่แว่นตาดำ และตรวจความผิดปกติ สำหรับ one-class classification [25], [26]

2.7 การเรียนรู้ของต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้โดยจำแนกประเภท (classification) ข้อมูลออกเป็นคลาส (class) ต่าง ๆ โดยใช้คุณลักษณะ (attribute) ของข้อมูลในการจำแนกประเภท ต้นไม้ตัดสินใจ (Decision Tree) ถูกพัฒนาโดย Ross Quinlan ในชื่อ ID3 algorithm ในปี ค.ศ.1986 [27] พัฒนาต่อเป็น C4.5 ในปี ค.ศ.1993 [28] ต้นไม้ตัดสินใจมีข้อจำกัดคือ ไม่สามารถสร้างต้นไม้ตัดสินใจที่ดีที่สุดให้ครอบคลุมทุกกรณีได้ และไม่สามารถทำการทดสอบแบบกล่องดำได้ (Black Box Testing) [29]

บทที่ 3
แนวคิดและวิธีการดำเนินการวิจัย

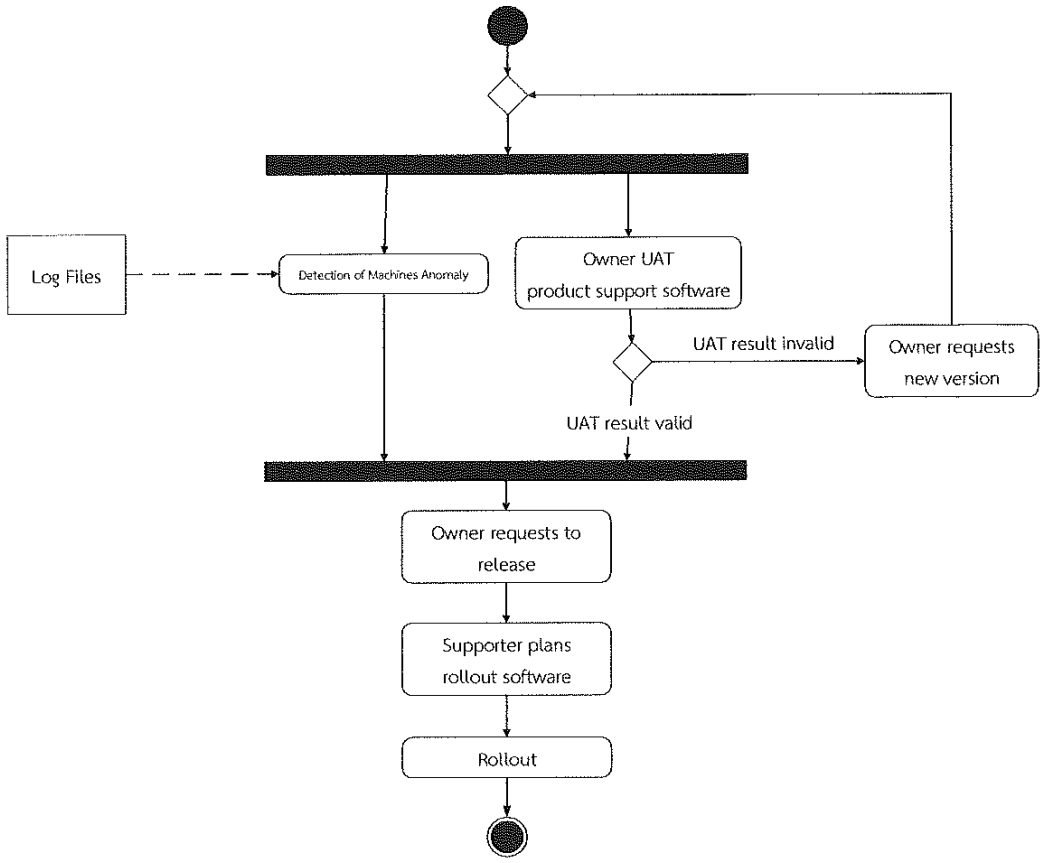
ตัวอย่างชนิดของปัญหาในโรงงานทั่วไป ได้แก่ เครื่องจักรเสียหาย ปัญหาคุณภาพ การวางแผนการผลิตไม่ดี และปัญหาด้านทักษะการทำงาน [30] เมื่อนำมาพิจารณาแล้วปัญหาแต่ละปัญหาก็จะมีอาการที่เหมือนกันบ้าง ต่างกันบ้าง คล้าย ๆ กันบ้าง หากนำประสบการณ์ของผู้ปฏิบัติงานแก้ไขปัญหามาทำการวิเคราะห์จะสามารถระบุได้ถึง 33 ปัญหา โดยวิทยานิพนธ์ฉบับนี้เสนอแนวคิดในการระบุ 5 ปัญหาซึ่งเกี่ยวข้องกับการประกอบชุดประกอบหัวอ่าน / เขียนสำเร็จ ได้แก่ ปัญหาพนักงานไม่ได้ลงทะเบียนหมายเลขเครื่องจักรเข้าไปในระบบ ปัญหาพนักงานลงทะเบียนเครื่องจักรผิดเงื่อนไขซอฟต์แวร์ที่เครื่องบริการถูกปรับปรุง ซอฟต์แวร์ที่เครื่องจักรถูกปรับปรุง และใบรับรองทักษะการทำงานของพนักงานที่ควบคุมเครื่องจักรหมดอายุ โดยรูปแบบการทำงานในปัจจุบันยังไม่มีการตรวจสอบปัญหาทั้ง 5 ปัญหา ในช่วงปรับปรุงซอฟต์แวร์ดังภาพที่ 9

5462082218



ภาพที่ 9 รูปแบบการทำงานในปัจจุบัน

แนวคิดที่นำเสนอคือเพิ่มการอ่านแฟ้มล็อกขณะปรับปรุงซอฟต์แวร์ เพื่อตรวจสอบทั้ง 5 ปัญหา ในช่วงปรับปรุงซอฟต์แวร์ดังภาพที่ 10

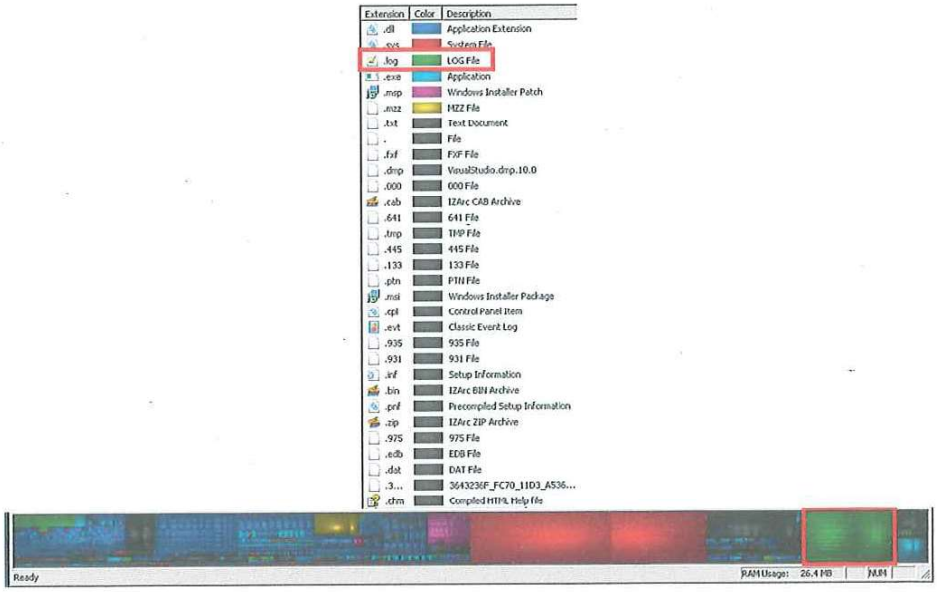


ภาพที่ 10 รูปแบบการทำงานเพิ่มเติมจากการอ่านแฟ้มล็อก

จากการสำรวจความจุแฟ้มล็อกของเครื่องบริการจำนวน 342 เครื่องที่สนับสนุนการผลิต เมื่อวันที่ 31 มกราคม ค.ศ.2015 และมีจำนวนแฟ้มถึง 263,850 แฟ้ม และประมาณความจุต่อ 7 วัน เป็นจำนวน 657.1595 กิกะไบต์

นำแนวคิดจากบทความของ Stephen G. Eick at el. (1994) มาทดลองทำให้แฟ้มล็อกจากเครื่องบริการที่ใช้ในการทดลองนี้มาสร้างจินตทัศน์ดังภาพที่ 11 แสดงให้เห็นความสัมพันธ์ของแฟ้มล็อกบนฮาร์ดดิสก์

545208218



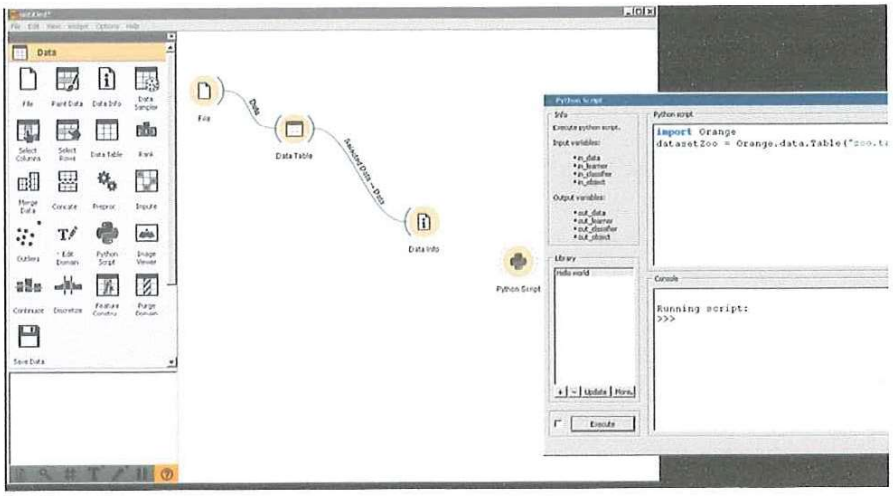
ภาพที่ 11 การสำรวจเครื่องบริการที่เลือกมาใช้งานวิจัยนี้

5462082218

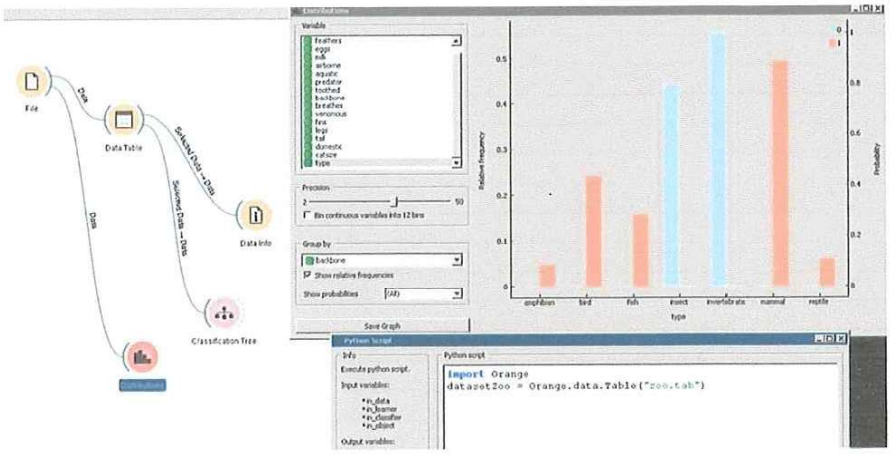
3.1 เครื่องมือที่เกี่ยวข้อง

3.1.1 Orange Canvas

Orange Canvas [31-34] อันเป็นชุดเครื่องมือสำหรับวิเคราะห์ข้อมูลด้วยการเรียนรู้ของเครื่องและการทำเหมืองข้อมูลพร้อมความเป็นโอเพนซอร์สพัฒนาจากโปรแกรมภาษา C++ ในปี ค.ศ.1996 โดยมหาวิทยาลัยสโลวีเนียและสถาบันเทคโนโลยีลูบลิยานาแล้วถูกนำมาต่อยอดด้วยภาษาไพทอนจนถึงปัจจุบัน ซึ่งได้กลายเป็นเครื่องมือหนึ่งที่กูเกิ้ลสนับสนุนให้นิสิตนักศึกษามากกว่า 10,900 คนจาก 103 ประเทศทั่วโลกเข้ามาพัฒนาต่อภายใต้โครงการ “Google Summer of Code” หรือ “GSoC” [35] ด้วยชื่อว่า “Orange at Google Summer of Code 2016” [36] โดยมีตัวอย่างการใช้งานดังภาพที่ 12 และภาพที่ 13



ภาพที่ 12 ตัวอย่างการนำชุดข้อมูลของสวนสัตว์เข้าโปรแกรม Orange Canvas



ภาพที่ 13 ตัวอย่างการแยกกลุ่มข้อมูลและสร้างมโนภาพด้วยโปรแกรม Orange Canvas

5462092218

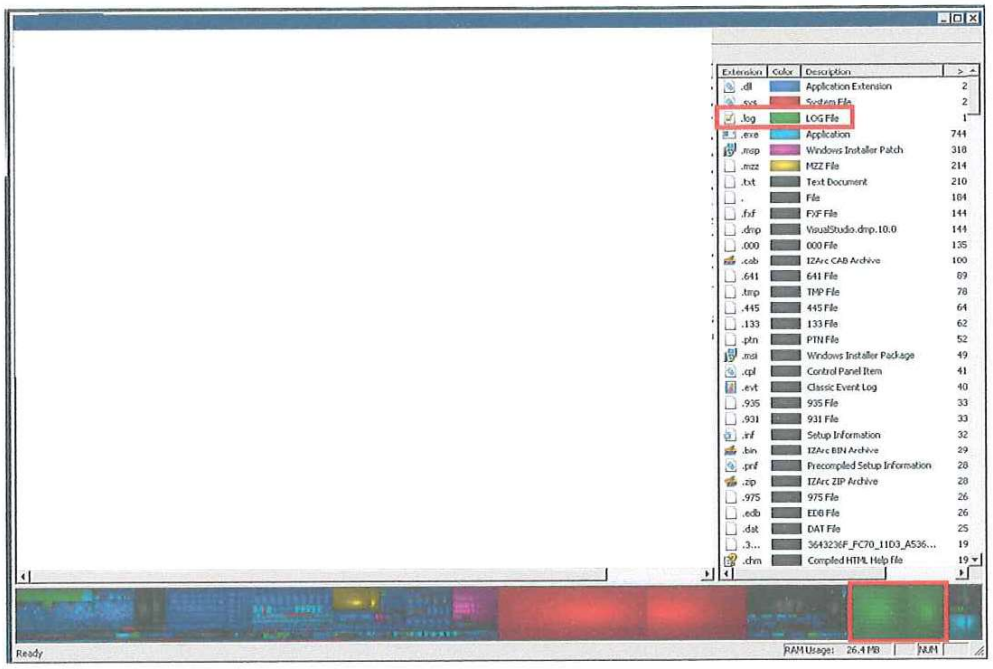
3.2 ข้อมูลเบื้องต้น (Preliminary Data)

ข้อมูลการสำรวจความจุแฟ้มล็อกของเครื่องบริการทุก ๆ เครื่อง เมื่อวันที่ 31 มกราคม ค.ศ.2015

ตารางที่ 1 ผลการสำรวจความจุของแฟ้มล็อก เมื่อวันที่ 31 มกราคม ค.ศ.2015

Size of Flat Files	Number of host = 342
	Number of files = 263,850 files
	Sizes = 657.1595 GB per 7 days

แฟ้มล็อกที่สนใจนำมาจากเครื่องบริการในช่วงการทดสอบซอฟต์แวร์ขั้นสุดท้าย (User Acceptance Testing หรือ UAT) ของการประกอบหัวอ่าน / เขียนสำเร็จ โดยสำรวจโครงสร้างแฟ้มข้อมูลในเครื่องบริการ เมื่อวันที่ 24 กันยายน ค.ศ.2015 เวลา 13:51:44 ดังภาพที่ 14



ภาพที่ 14 การสำรวจเครื่องบริการที่บันทึกแฟ้มล็อก

3462078

โดยซอฟต์แวร์ที่สนับสนุนกระบวนการผลิตดังกล่าวทำงานสนับสนุนสายการผลิตในคลีนรูม (Clean Room) คือห้องสะอาดที่จำกัดฝุ่นไม่ให้มีขนาดใหญ่เกินกว่าที่กำหนด โดยแฟ้มล็อกมีลักษณะที่เก็บ (Log file path) อยู่ที่ "C:\P1\P2\P3\LogFile\" ผ่านการสำรวจรายละเอียดในแฟ้มล็อกเมื่อวันที่ 19 ตุลาคม ค.ศ.2015 เวลา 16:38:06 - 16:50:39 น. ในลักษณะกระแสข้อมูล (Steaming) ผ่านกลุ่มทำงาน (Batch) ทุก ๆ ชั่วโมง โดยจัดเก็บใน Folder ประกอบด้วยตัวเลข 6 ตัวคือ เดือน วัน และชั่วโมง คือ mmHHdd

ตัวอย่างรูปแบบ คือ Timestamp เว้นด้วยเครื่องหมาย : แต่ละ Transaction จะเริ่มและจบ ดังนี้ Timestamp : Status of transaction: Message Type, LINE Type, Serial Number, Line number, ,,,,...,

โครงสร้างแฟ้มล็อกในแต่ละรายการเปลี่ยนแปลงเป็นดังนี้

Timestamp : R: Message number and details

Timestamp : A: Message number and details

Timestamp : A: Message number and details

.

.

.

Timestamp : A: Message number and details

Timestamp : S: Message number and details

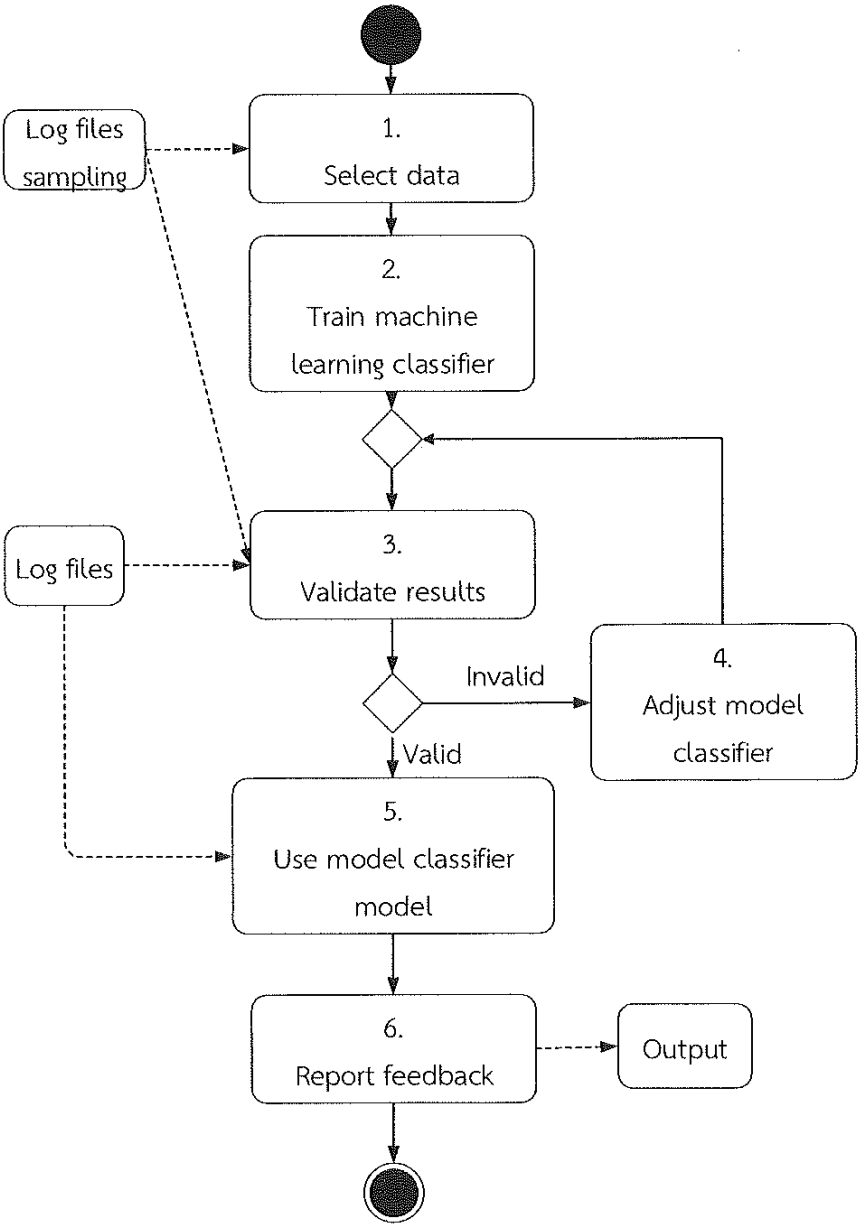
Timestamp : Message, Response Time: time seconds

แฟ้มล๊อกว่าในแต่ละรายการเปลี่ยนแปลง โดยมีบรรทัดที่มีอักษร R: เป็นจุดเริ่มต้น ตามด้วยบรรทัดที่มี A: ต่อด้วยบรรดทัดที่มี S: และและบรรทัดปิดท้ายที่มี Response Time:

มีหลายสาเหตุที่แสดงถึงความผิดปกติของเครื่องจักรในสายการผลิต ที่มีข้อผิดพลาดบางอย่างที่เกิดจากมนุษย์ เช่น พนักงานป้อนรหัสเครื่องที่ไม่ถูกต้องหรือสิทธิ์การทำงานหมดอายุ ข้อผิดพลาดส่วนมากจะเกิดจากการเปลี่ยนแปลงในการกำหนดค่าซอฟต์แวร์บนเครื่องบริการและเครื่องจักรควบคุมการผลิต ตัวอย่างเช่น การปรับปรุงซอฟต์แวร์และเฟิร์มแวร์ ข้อผิดพลาดอีกอย่างคือข้อผิดพลาดทางกายภาพ เป็นต้นว่าไฟฟ้าดับและความล้มเหลวทางกลของเครื่องจักรในสายการผลิต ข้อผิดพลาดเหล่านี้สามารถตรวจพบได้ง่ายดาย โดยพนักงานที่เกี่ยวข้อง ส่วนใหญ่ของข้อผิดพลาดที่สำคัญมักถูกส่งออกมาจากซอฟต์แวร์ที่ไม่สามารถเห็นและอ่านได้ด้วยตา เป็นผลให้ไม่สามารถแก้ไขได้ทันที

จะยอดเยี่ยมเพียงใดหากนำโปรแกรมคอมพิวเตอร์เข้ามาทำงานนี้แทนมนุษย์ โดยอาศัยประสบการณ์ของมนุษย์ทำงานแทนผ่านการเรียนรู้ของเครื่อง งานวิจัยนี้จึงได้นำเสนอแนวคิดและวิธีการทำงานดังภาพที่ 15

546208218



ภาพที่ 15 กิจกรรมแนวคิดในการดำเนินงานวิจัย

ในแผนภาพกิจกรรมเริ่มจากการเตรียมข้อมูลในข้อที่ 1 โดยดึงข้อมูลจากกลุ่มตัวอย่างแฟ้มล็อกที่เตรียมไว้ แล้วนำมาฝึกสอนการเรียนรู้ของเครื่องในข้อที่ 2 เพื่อแยกรายการเปลี่ยนแปลงที่ปกติและผิดปกติ แล้วนำมาประเมินผลการเรียนรู้ว่ารับได้เพียงใด หากผลการประเมินรับไม่ได้ต้องนำไปปรับแต่งพารามิเตอร์ที่เกี่ยวข้องในข้อที่ 4 แต่หากผลยอมรับได้ก็นำมาใช้เป็นต้นแบบในข้อที่ 5 และนำเสนอผลลัพธ์ในข้อที่ 6 เพื่อออกเป็นรายงานสรุปเพื่อนำส่งผู้พัฒนาซอฟต์แวร์

54620821

บทที่ 4
การดำเนินงานวิจัย

บทนี้กล่าวถึงรายละเอียดในการทำวิจัย ขั้นแรกจะประมวลผลแฟ้มล็อกก่อน เพื่อแปลงให้เป็นข้อมูลที่เหมาะสมแล้วใช้การเรียนรู้ของเครื่อง (Machine Learning) เปรียบเทียบกันทั้งสี่วิธี ระหว่างซัพพอร์ตเวกเตอร์แมชชีน เคเอ็นเอ็น นาอีฟเบส์ และต้นไม้ตัดสินใจ จากนั้นศึกษาปัจจัยที่ทำให้เกิดข้อผิดพลาดอย่างละเอียดขึ้นโดยใช้ต้นไม้ตัดสินใจ โดยการเรียนจากคุณลักษณะที่ละเอียดขึ้นถึง 36 ปัจจัย

4.1 การประมวลผลก่อน (Preprocessing)

แฟ้มล็อกก่อนนำเข้าการเรียนรู้ของเครื่องจะถูกประมวลผลก่อนด้วยโปรแกรมภาษาไพทอนเพื่อสกัดคุณลักษณะต่าง ๆ ก่อนนำเข้าการเรียนรู้ของเครื่อง 11 คุณลักษณะขาเข้า และอีกหนึ่งคุณลักษณะซึ่งเป็นผลลัพธ์ คุณลักษณะขาเข้า ได้แก่

5462028218

คุณลักษณะที่ 1 หมายเลขข้อความเป็นข้อมูลวิยุต (Discrete) อันสะท้อนถึงชนิดการขอรับบริการจากเครื่องจักร

คุณลักษณะที่ 2 จำนวนคำว่า “Error” ในแต่ละรายการการเปลี่ยนแปลง เป็นข้อมูลต่อเนื่องสะท้อนถึงความถี่ของการเกิดความผิดพลาด

คุณลักษณะที่ 3 ลำดับของสถานะภายในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต หากลำดับเป็นการรับข้อความ การประมวลผล การส่งกลับข้อความ และการบันทึกเวลาการตอบสนองหากเรียงลำดับตามนี้ให้ค่าเป็น 1 แต่ถ้าไม่ใช่ให้เป็น 0

คุณลักษณะที่ 4 เวลาการเริ่มต้นของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง มีหน่วยเป็นวินาที โดยใช้วันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น..เป็นจุดอ้างอิงที่เริ่มต้นด้วยศูนย์ สะท้อนถึงเวลาเริ่มต้นของการขอรับบริการจากเครื่องจักร

คุณลักษณะที่ 5 เวลาสุดท้ายของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง มีหน่วยเป็นวินาทีนับจากวันที่ 1 มกราคาม ค.ศ.1970 เวลา 7:00 น. เป็นจุดอ้างอิงที่เริ่มต้นด้วยศูนย์ สะท้อนถึงเวลาที่เครื่องจักรได้รับผลของการบริการ

คุณลักษณะที่ 6 เวลาการตอบสนองแต่ละรายการการเปลี่ยนแปลงมีหน่วยเป็นวินาทีและเป็นข้อมูลต่อเนื่อง สะท้อนถึงเวลาที่เครื่องจักรได้รับการตอบสนองในแต่ละรายการการเปลี่ยนแปลง

คุณลักษณะที่ 7 เวลาของแฟ้มล็อกถูกปรับปรุงครั้งสุดท้ายแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่องมีหน่วยเป็นวินาทีนับจากวันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น..เป็นจุดอ้างอิงที่เริ่มต้นด้วยศูนย์ สะท้อนถึงการบันทึกแฟ้มจากเครื่องบริการในครั้งสุดท้ายของแต่ละชั่วโมง

คุณลักษณะที่ 8 จำนวนบรรทัดของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง มีหน่วยเป็นบรรทัด

คุณลักษณะที่ 9 จำนวนคำว่า “No Data” ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่องสะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะที่ 10 จำนวนคำว่า “Cert” ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่องสะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะที่ 11 จำนวนคำว่า “Expire” ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่องสะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะผลลัพธ์เพื่อใช้ในการสอนเป็นข้อมูลวิยุตมีเพียงสองค่าคือ Normal แสดงถึงรายการการเปลี่ยนแปลงปกติ และ Abnormal แสดงถึงรายการการเปลี่ยนแปลงที่ผิดปกติ

โดยทำการทดลองเปรียบเทียบการเรียนรู้ของเครื่อง 4 วิธีการ ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตันใจ เคเอ็นเอ็น และนาอีฟเบส์ โดยนำแฟ้มล็อกตั้งแต่วันที่ 12 ตุลาคม ค.ศ.2015 ถึงวันที่ 13 มกราคม ค.ศ.2016 ทำการประมวลผลก่อนได้ผลจำนวน 104,582 ระเบียน โดยซัพพอร์ตเวกเตอร์แมชชีนและต้นไม้ตันใจดีสุด

ตารางที่ 2 คำอธิบายข้อมูลหลังการประมวลผลก่อน

คุณลักษณะ	ชนิดของคุณลักษณะ	ตัวอย่าง	การสะท้อนถึงปัญหา
Message Number	วิยุต	80003	
Error Count	แบบต่อเนื่อง (Continuous)	0	ซอฟต์แวร์ที่เครื่องบริการถูกปรับปรุงหรือซอฟต์แวร์ที่เครื่องจักรถูกปรับปรุง
Sequence Pattern Transaction	วิยุต	1	
Start Transaction	แบบต่อเนื่อง	1444617655.044	
Stop Transaction	แบบต่อเนื่อง	1444617656.841	
Response Time	แบบต่อเนื่อง	1.7968635	
Modify Log file Timestamp	แบบต่อเนื่อง	1444620430	
Lines of Transaction	แบบต่อเนื่อง	7	
No Data Count	แบบต่อเนื่อง	2	ปัญหาพนักงานลงทะเบียนหมายเลขเครื่องจักรผิดเงื่อนไขหรือปัญหาพนักงานไม่ได้ลงทะเบียนหมายเลขเครื่องจักรเข้าไปในระบบ
Cert Count	แบบต่อเนื่อง	0	ใบรับรองทักษะการทำงานของพนักงานที่คุมเครื่องจักรหมดอายุ
Expire Count	แบบต่อเนื่อง	0	
Transaction Type	วิยุต	Normal	ผลลัพธ์

	Message Number	Error Count	Sequence Pattern Transaction	Start Transaction	Stop Transaction	Response Time	Modify Logfile Timestamp	Lines of Transaction	Cert Count	Expire Count	No Data Count	Transaction Type
2	d	c	d	c	c	c	c	c	c	c	d	
3											class	
4	80003	0	1	1444617655.044	1444617656.841	1.7968635	1444620430	7	2	0	0	Abnormal
5	80003	0	1	1444617667.701	1444617669.638	1.9374876	1444620430	35	0	0	0	Normal
6	80003	0	1	1444617682.044	1444617683.529	1.4843655	1444620430	6	2	0	0	Abnormal
7	80003	0	1	1444617694.825	1444617696.185	1.3593663	1444620430	6	2	0	0	Abnormal
8	80003	0	1	1444617697.138	1444617698.575	1.4374908	1444620430	6	2	0	0	Abnormal
9	80003	0	1	1444617703.747	1444617704.341	0.5937462	1444620430	6	2	0	0	Abnormal
10	80003	0	1	1444617711.06	1444617711.497	0.4374972	1444620430	6	2	0	0	Abnormal
11	80003	0	1	1444617759.231	1444617760.856	1.6249896	1444620430	6	2	0	0	Abnormal
12	80003	0	1	1444617766.622	1444617767.184	0.5624964	1444620430	6	2	0	0	Abnormal
13	80003	0	1	1444617780.419	1444617781.997	1.5781149	1444620430	6	2	0	0	Abnormal

ภาพที่ 16 ตัวอย่างข้อมูลที่ถูกประมวลผลก่อน

546208218

4.2 การสร้างแบบจำลองและนำแบบจำลองที่ได้ไปใช้ (Model Construction and Model Usage)

หลังการประมวลผลก่อนทำให้ได้ชุดข้อมูลสอน (training data set) โดยแต่ละระเบียนจะมีคุณลักษณะหนึ่งซึ่งเป็นผลลัพธ์บอกค่าไว้แล้วว่าเป็นระเบียนปกติหรือผิดปกติ แล้วสร้างแบบจำลองแสดงผลการเรียนรู้การจำแนกประเภทในรูปแบบ 4 วิธี คือ ซัพพอร์ตเวกเตอร์แมชชีน เคเอ็นเอ็น การเรียนรู้แบบเบย์ และต้นไม้ตัดสินใจ

4.3 การประเมินค่า (Evaluation)

ผลลัพธ์ของการฝึกสอนโดยการเรียนรู้ของเครื่องด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เคเอ็นเอ็น การเรียนรู้แบบเบย์ และต้นไม้ตัดสินใจ การประเมินค่าการสอน (Train) และทดสอบ (Test) โดยใช้ 10 - Fold Validation เป็นการแบ่งข้อมูลออกเป็น 10 ชุดโดยแต่ละชุดแบ่งออกเป็น 10 ส่วน 9 ส่วนเพื่อสอน และอีก 1 ส่วนเพื่อทดสอบ ทำเช่นนี้ 10 รอบ แล้วคำนวณค่าความแม่นยำจากอัตราส่วนระหว่างจำนวนตัวอย่างในชุดข้อมูลทดสอบที่ทำนายกลุ่มได้อย่างถูกต้องทั้งหมด 10 รอบกับจำนวนตัวอย่างทั้งหมด ซึ่งได้ผลดังตารางที่ 3 ตารางที่ 4 และตารางที่ 5

546208218

ตารางที่ 3 เปรียบเทียบประสิทธิภาพของการเรียนรู้ของเครื่อง 4 วิธี

Overall score from all data				
	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000
k-NN	0.999	1.000	1.000	1.000
Naïve Bayes	0.990	0.995	1.000	0.990
Decision Tree	1.000	1.000	1.000	1.000

ตารางที่ 3 แสดงถึงผลการเรียนด้วยซัพพอร์ตเวกเตอร์แมชชีนและต้นไม้ตันใจได้ผลที่ดีที่สุด ลำดับถัดมาเป็นเคเอ็นเอ็น และลำดับสุดท้ายคือนาอีฟเบส์ ตามลำดับ

ตารางที่ 4 เปรียบเทียบประสิทธิภาพของการเรียนรู้ข้อมูลที่ผิดปกติ 4 วิธีการเรียนรู้ของเครื่อง

Score from Anomaly group				
	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000
k-NN	0.999	0.751	0.890	0.650
Naïve Bayes	0.990	0.195	0.109	0.956
Decision Tree	1.000	0.996	1.000	0.993

ตารางที่ 4 แสดงถึงผลการเรียนข้อมูลที่ผิดปกติ ออกมาว่าซัพพอร์ตเวกเตอร์แมชชีนได้ผลดีที่สุด ลำดับถัดมาเป็นต้นไม้ตัดสินใจ ตามมาด้วยเคเอ็นเอ็น และนาอีฟเบส์ ตามลำดับ

ตารางที่ 5 เปรียบเทียบประสิทธิภาพของการเรียนรู้ข้อมูลปกติ 4 วิธีการเรียนรู้ของเครื่อง

Score from Normal group				
	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000
k-NN	0.999	1.000	1.000	1.000
Naïve Bayes	0.990	0.995	1.000	0.990
Decision Tree	1.000	1.000	1.000	1.000

5462028218

ตารางที่ 5 แสดงถึงผลการเรียนข้อมูลปกติ ออกมาว่าซัพพอร์ตเวกเตอร์แมชชีนและต้นไม้ตัดสินใจได้ผลดีที่สุด ลำดับถัดมาเป็นเคเอ็นเอ็น และลำดับสุดท้ายคือนาอีฟเบส์ ตามลำดับ

4.4 การเรียนรู้ด้วยต้นไม้ตัดสินใจอย่างละเอียด

จากตารางที่ 3 ตารางที่ 4 และตารางที่ 5 บอกถึงผลที่เรียนรู้ได้ทั้ง วิธี ในการวัดประสิทธิภาพ 4 ความแม่นยำในการจำแนกข้อมูลที่ผิดปกติ โดยมีซัพพอร์ตเวกเตอร์แมชชีนดีสุด ลองลงมาเป็นต้นไม้วิธีมีประสิทธิภาพความ 4 ตัดสินใจ แล้วจึงเป็นเคเอ็นเอ็น และการเรียนรู้เบส์อย่างง่าย กล่าวคือทั้งแม่นยำในการจำแนกดีโดยไม่มีนัยสำคัญเนื่องจากข้อมูลเสียดุลอันเป็นธรรมชาติของข้อมูลจากโรงงานผลิตระดับหกซิกมา คือในการผลิตหนึ่งล้านชิ้นจะมีของเสียไม่เกิน 3.4 ชิ้น

เมื่อเทียบกับวิธีที่ใช้ในงานวิจัยของ Stephen G. Tick et al. (1994) เพื่อการสร้างเครื่องมือให้มนุษย์ดูได้ง่ายขึ้น จึงยังต้องอาศัยมนุษย์ในการวิเคราะห์อยู่ แต่งานวิทยานิพนธ์นี้เน้นการสร้าง

เครื่องมือเพื่อให้การตรวจสอบทำได้อย่างเป็นอัตโนมัติอันช่วยลดความยุ่งยากและลดเวลาการทำงานของมนุษย์ได้ คือทำงานบางงานทดแทนมนุษย์ได้ เมื่อเทียบกับผลงานที่นำแฟ้มล็อกมาประมวลผลอย่างอัตโนมัติของ James H. Andrews (1998) พยายามทำให้เป็นการวิเคราะห์อัตโนมัติกับระบบซอฟต์แวร์ขนาดใหญ่ โดยพัฒนาภาษาขึ้นมาใหม่ชื่อว่า LFAL (Log Files Analysis Language) [37] ซึ่งทำให้ง่ายในการวิเคราะห์และจำลองรูปแบบเป็น State Machine Diagram โดยไม่ต้องสร้างกรณีทดสอบด้วยมนุษย์ มีความยืดหยุ่นใช้ได้หลายภาษา แต่มีข้อจำกัดว่านักพัฒนาและนักทดสอบต้องมาเรียนรู้ภาษา LFAL และยังไม่สามารถทำเป็น Formal Verification ได้ ต่างจากงานวิทยานิพนธ์ฉบับนี้คือใช้การเรียนรู้ของเครื่องทำการวิเคราะห์แฟ้มล็อกอัตโนมัติ โดยที่นักพัฒนาและนักทดสอบไม่จำเป็นต้องมาเรียนรู้ภาษาใหม่สำหรับการวิเคราะห์ผล สำหรับการเขียนโปรแกรมได้นำโปรแกรมสคริปภาษาไพทอนในการประมวลผลก่อน ซึ่งเป็นภาษาที่ง่ายต่อการเรียนรู้เช่นกัน ดังนั้นในวิทยานิพนธ์ฉบับจึงนำสนอผลงานที่เป็นอัตโนมัติ ง่าย และมีความชัดเจนในการวิเคราะห์มากขึ้น

จากการทดลองเพื่อตอบคำถามว่าคุณลักษณะใดในแฟ้มล็อกมีความสัมพันธ์กับความผิดปกติของเครื่องจักร เพราะปัญญาประดิษฐ์ทำให้ทราบความสัมพันธ์ของคุณลักษณะต่าง ๆ ชัดเจน ดังนั้นการสกัดคุณลักษณะต่าง ๆ ในแฟ้มล็อกด้วยการนำประสบการณ์จากวิศวกรที่แก้ปัญหามา 4 ปี โดยอาศัยแฟ้มล็อกเป็นหลักฐานในสืบเสาะหาสาเหตุและผลกระทบของปัญหาที่เกิดขึ้น โปรแกรมภาษาไพตอนจึงถูกปรับปรุงเพื่อการประมวลข้อมูลก่อนจนได้คุณลักษณะขาเข้าจนได้จำนวน 29 คุณลักษณะ ประกอบด้วยข้อมูลวิยุตจำนวน 20 คุณลักษณะ และข้อมูลต่อเนื่องจำนวน 9 คุณลักษณะ เพิ่มเติมข้อมูลเมตตาเดตาที่เป็นประโยชน์ในวิเคราะห์ข้อมูลภายหลังได้อีกจำนวน 7 อย่าง รวมทั้งสิ้น 36 คุณลักษณะ ได้แก่

คุณลักษณะที่ 1 ชื่อแฟ้มล็อกเป็นข้อมูลเมตาเดตา อันสะท้อนถึงแหล่งข้อมูลดิบ

คุณลักษณะที่ 2 วิถีของแฟ้มล็อกเป็นข้อมูลเมตาเดตา อันสะท้อนถึงตำแหน่งที่เก็บแหล่งข้อมูลดิบ

คุณลักษณะที่ 3 ขนาดของแฟ้มล็อกเป็นข้อมูลต่อเนื่องมีหน่วยเป็นไบต์ อันสะท้อนถึงปริมาณข้อมูลที่เก็บในแต่ละแฟ้มล็อก

คุณลักษณะที่ 4 เวลาที่สร้างแฟ้มล็อกเป็นข้อมูลเมตาเดตา อันสะท้อนเวลาเริ่มต้นที่แฟ้มนี้ถูกเขียน

คุณลักษณะที่ 5 เวลาที่สร้างแฟ้มล็อกเป็นข้อมูลต่อเนื่องมีหน่วยเป็นวินาที นับจากวันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น. อันสะท้อนเวลาเริ่มต้นที่แฟ้มนี้ถูกเขียน

คุณลักษณะที่ 6 เวลาที่ปรับปรุงแฟ้มล็อกเป็นข้อมูลเมตาเดตา อันสะท้อนเวลาล่าสุดที่แฟ้มนี้ถูกปรับปรุง

3462082118

คุณลักษณะที่ 7 เวลาที่ปรับปรุงแฟ้มล็อกเป็นข้อมูลต่อเนื่องมีหน่วยเป็นวินาที นับจากวันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น. อันสะท้อนเวลาล่าสุดที่แฟ้มนี้ถูกปรับปรุง

คุณลักษณะที่ 8 หมายเลขข้อความเป็นข้อมูลวิยุต อันสะท้อนถึงชนิดการขอรับบริการจากเครื่องจักร

คุณลักษณะที่ 9 จำนวนคำว่า "Error" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของการเกิดความผิดพลาด

คุณลักษณะที่ 10 ลำดับของสถานะภายในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต โดยมีลำดับเป็นการรับข้อความ การประมวลผล การส่งกลับข้อความ และการบันทึกเวลาการตอบสนอง หากเรียงลำดับตามนี้ให้ค่าเป็น 1 แต่ถ้าไม่ใช่ให้เป็น 0

คุณลักษณะที่ 11 จำนวนสถานะรับข้อมูลจากเครื่องจักรแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 12 จำนวนสถานะประมวลผลข้อมูลแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 13 จำนวนสถานะส่งข้อมูลกลับไปยังเครื่องจักรแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 14 จำนวนสถานะที่คำนวณเวลาการตอบสนองแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 15 เวลาการเริ่มต้นของแต่ละรายการการเปลี่ยนแปลงเป็นเมตาเดตา สะท้อนถึงเวลาเริ่มต้นของการขอรับบริการของเครื่องจักร

คุณลักษณะที่ 16 เวลาการเริ่มต้นของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง มีหน่วยเป็นวินาทีนับจากวันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น. สะท้อนถึงเวลาเริ่มต้นของการขอรับบริการของเครื่องจักร

คุณลักษณะที่ 17 เวลาการสิ้นสุดของแต่ละรายการการเปลี่ยนแปลงเป็นเมตาเดตา สะท้อนถึงเวลาที่เครื่องจักรได้รับผลลัพธ์

คุณลักษณะที่ 18 เวลาการสิ้นสุดของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง มีหน่วยเป็นวินาทีนับจากวันที่ 1 มกราคม ค.ศ.1970 เวลา 7:00 น. สะท้อนถึงเวลาที่เครื่องจักรได้รับผลลัพธ์

คุณลักษณะที่ 19 ความต่างกันของคุณลักษณะที่ 16 และ 18 มีหน่วยเป็นวินาทีเป็นข้อมูลต่อเนื่อง สะท้อนจำนวนเวลาที่ใช้ในแต่ละรายการการเปลี่ยนแปลง

คุณลักษณะที่ 20 ความต่างกันของคุณลักษณะที่ 7 และ 18 มีหน่วยเป็นวินาทีเป็นข้อมูลต่อเนื่อง สะท้อนความต่างกันของเวลาล่าสุดในแต่ละรายการการเปลี่ยนและเวลาที่ปรับปรุงแฟ้มล็อก

54608218

คุณลักษณะที่ 21 ความต่างกันของเวลาการตอบสนองของแต่ละรายการการเปลี่ยนแปลง มีหน่วยเป็นวินาทีเป็นข้อมูลต่อเนื่อง สะท้อนถึงเวลาที่เครื่องจักรได้รับการตอบสนองในแต่ละรายการการเปลี่ยนแปลง

คุณลักษณะที่ 22 จำนวนบรรทัดของแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 23 จำนวนคำว่า "No Data" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะที่ 24 จำนวนคำว่า "Cert" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะที่ 25 จำนวนคำว่า "Expire" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของความสำคัญผิด

คุณลักษณะที่ 26 จำนวนคำว่า "Fail" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของการเกิดความผิดพลาด

คุณลักษณะที่ 27 จำนวนคำว่า "Abort" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของการเกิดการขัดจังหวะ

คุณลักษณะที่ 28 จำนวนคำว่า "Block" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของการถูกควบคุมคุณภาพชิ้นงาน

คุณลักษณะที่ 29 จำนวนคำว่า "Validate" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของการถูกควบคุมคุณภาพชิ้นงาน

คุณลักษณะที่ 30 จำนวนคำว่า "TX" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของความเกี่ยวข้องกับฐานข้อมูลการสื่อสัญญาณ (Transmission)

คุณลักษณะที่ 31 จำนวนคำว่า "BR" ในแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต สะท้อนถึงความถี่ของความเกี่ยวข้องกับฐานข้อมูลบรานช์ (Branch)

คุณลักษณะที่ 32 ชั่วโมงที่เกิดรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 33 นาทีที่เกิดรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

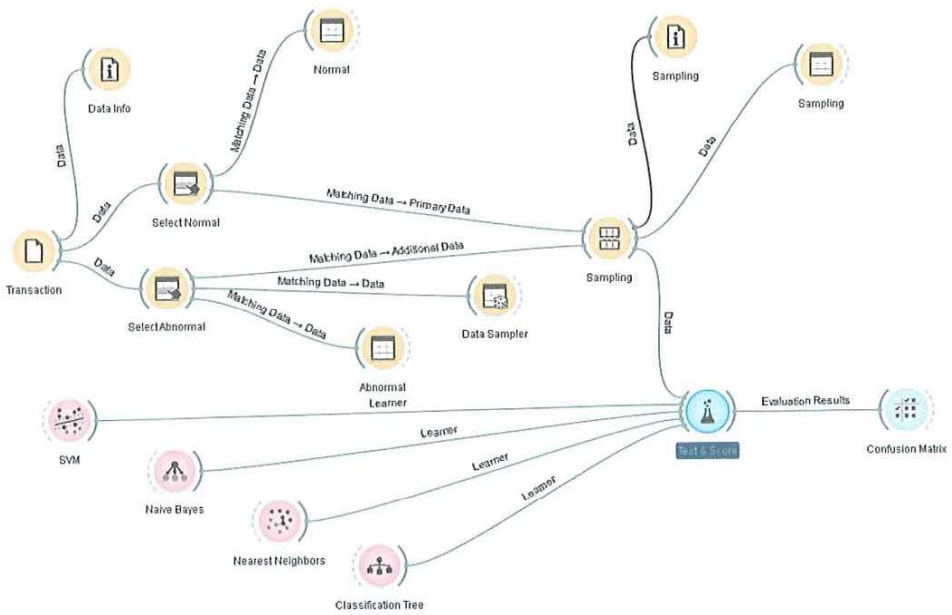
คุณลักษณะที่ 34 วินาทีที่เกิดรายการการเปลี่ยนแปลงเป็นข้อมูลต่อเนื่อง

คุณลักษณะที่ 35 หมายเลขของการตอบสนองแต่ละรายการการเปลี่ยนแปลงเป็นข้อมูลวิยุต

คุณลักษณะที่ 36 ข้อความที่ตอบสนองแต่ละรายการการเปลี่ยนแปลงเป็นเมตาเดตา

คุณลักษณะผลลัพธ์เพื่อใช้ในการสอนเป็นข้อมูลวิยุตมีเพียงสองค่าคือ Normal แสดงถึงรายการการเปลี่ยนแปลงปกติ และ Abnormal แสดงถึงรายการการเปลี่ยนแปลงที่ผิดปกติ ดำเนินการทดลองดังภาพที่ 17

9460827 18



ภาพที่ 17 การทดลองเพื่อแยกกลุ่มข้อมูลปกติและผิดปกติโดย Orange Canvas

5462082218

ผลลัพธ์ของการฝึกสอนโดยการเรียนรู้ของเครื่องด้วยต้นไม้ตัดสินใจ การประเมินค่าการสอน (Train) และทดสอบ (Test) โดยใช้ -10 Fold Validation ได้ผลร้อยเปอร์เซ็นต์หรือดีเยี่ยมดังตารางที่ 6 ตารางที่ 7 และตารางที่ 8

ตารางที่ 6 ประสิทธิภาพของการเรียนรู้ของเครื่องด้วยต้นไม้ตัดสินใจ

Overall score from all data				
	CA	F1	Precision	Recall
Decision Tree	1.000	1.000	1.000	1.000

ตารางที่ 7 ประสิทธิภาพของการเรียนรู้ของเครื่องกับข้อมูลที่ผิดปกติด้วยต้นไม้ตัดสินใจ

Score from Anomaly group				
	CA	F1	Precision	Recall
Decision Tree	1.000	1.000	1.000	1.000

ตารางที่ 8 ประสิทธิภาพของการเรียนรู้ของเครื่องกับข้อมูลปกติด้วยต้นไม้ตัดสินใจ

Score from Normal group				
	CA	F1	Precision	Recall
Decision Tree	1.000	1.000	1.000	1.000

5462062718

เนื่องจากข้อมูลในงานจริงมีข้อมูลไม่สมบูรณ์ (incomplete) ข้อมูลรบกวน (noise) ข้อมูลไม่สอดคล้อง (inconsistent) จึงต้องมีการประมวลผลข้อมูลก่อน เพื่อดึงลักษณะสำคัญก่อนจะป้อนเข้าสู่โปรแกรมการเรียนรู้ของเครื่องต่อไป แฟ้มล็อกตั้งแต่วันที่ 12 ตุลาคม ค.ศ. 2014 ถึงวันที่ 13 มกราคม ค.ศ. 2015 จำนวน 69 แฟ้ม รวมขนาด 74.7 เมกะไบต์ ตัดเป็นระเบียนละหนึ่งเหตุการณ์ ข้อมูลหลังทำการข้อมูลประมวลผลก่อน ได้ข้อมูล 104,500 ระเบียน ประกอบด้วยลักษณะสำคัญทั้งหมด 29 อย่าง ประกอบด้วยข้อมูลวิยุต (Discrete) จำนวน 20 อย่าง และข้อมูลชนิดตัวเลขต่อเนื่อง จำนวน 9 อย่าง ใช้ในการวิเคราะห์สาเหตุของการผิลดพลาด โดยข้อมูลมีสองกลุ่มคือเป็นกลุ่มข้อมูลปกติ 104,381 ระเบียนและข้อมูลผิดปกติ 119 ระเบียน ซึ่งถูกระบุโดยผู้เชี่ยวชาญที่รับผิดชอบในการวินิจฉัยของสายการผลิต คุณลักษณะเหล่านี้สามารถครอบคลุมความผิดปกติที่หลากหลายแต่ไม่ใช่ทั้งหมด การศึกษาถูกจำกัดด้วยปริมาณข้อมูลผิดปกติที่หายากเพราะเกิดขึ้นน้อยในโรงงานผลิตระดับหกซิกมา นั่นคือมีของเสียไม่เกิน 3.4 ชิ้น ต่อการผลิตหนึ่งล้านชิ้น ยกตัวอย่างเช่น โรงงานผลิตฮาร์ดดิสก์

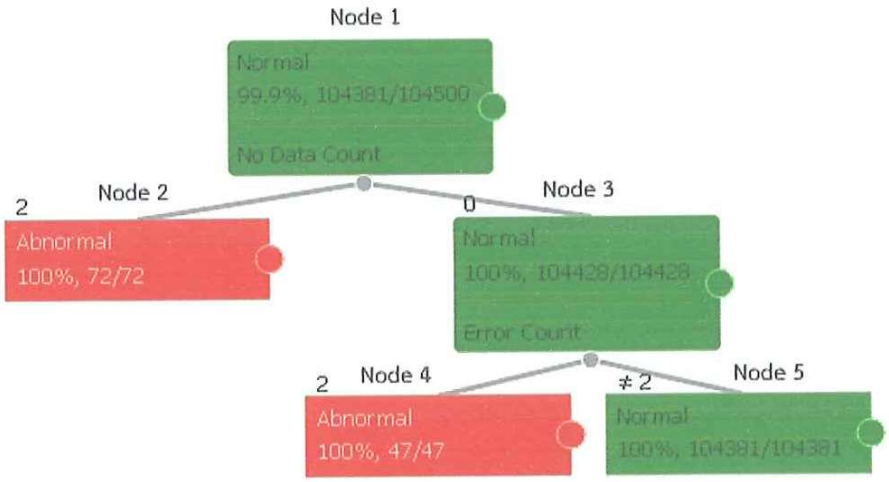
ลักษณะสำคัญประกอบด้วย กลุ่มใหญ่ ๆ กลุ่มที่หนึ่งคือการการบูรณาการข้อมูล 3(Data Integration) ได้แก่ การนับคำ ต่อหนึ่งหน่วยการเปลี่ยนแปลง (Transaction) จากการแจ้งความผิดปกติ ด้วยคำว่า “No Data” “Error” “R: ” “A: ” “S: ” “Response Time” “Cert” “Expire” “Fail” “Abort” “Block” “Validate” “TX” และ “BR” กลุ่มที่สองคือการแปลงข้อมูล (Data Transformation) โดยนำค่าต่าง ๆ ที่ตัดออกมาเป็นความหมายที่มนุษย์สามารถเข้าใจได้ ได้แก่ เวลาการตอบสนองต่อหนึ่งหน่วยการเปลี่ยนแปลง เวลาที่บันทึกไว้ขณะข้อความตอบกลับจากระบบสามารถนำมาแยกวิเคราะห์รายชั่วโมง รายนาที และรายวินาที ที่เกิดล็อกดังกล่าว ขนาดแฟ้มล็อก เวลาเริ่มต้นและเวลาสิ้นสุดของแต่ละรายการการเปลี่ยนแปลง เวลาที่ปรับปรุงแฟ้มล็อกครั้งล่าสุด

หมายเลขข้อความ ความถูกต้องของลำดับการบันทึก จำนวนบรรทัดต่อรายการเปลี่ยนแปลง เวลาเริ่มต้นที่เริ่มเขียนแฟ้มล็อก กลุ่มที่สามคือการลดข้อมูล (Data Reduction) เช่น ค่าการคำนวนช่วงเวลาของเวลาเริ่มต้นและเวลาสิ้นสุดของแต่ละหน่วยการเปลี่ยนแปลง และช่วงเวลาของเวลาสุดท้ายที่แฟ้มล็อกถูกปรับปรุงกับเวลาสุดท้ายของแต่ละการเปลี่ยนแปลง

โดยก่อนนำข้อมูลเข้าเครื่องมือ Orange Canvas ผู้วิจัยเขียนโปรแกรมประมวลผลก่อนด้วยภาษาไพทอนเพื่อสกัดสาระในข้อมูลจากแฟ้มล็อกออกมาเป็นแต่ละรายการการเปลี่ยนแปลง โดยแต่ละรายการการเปลี่ยนแปลงสะท้อนถึงเหตุการณ์หนึ่ง ๆ ด้วยวิธีการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจเพื่อเรียนรู้รูปแบบและแยกแยะข้อมูลที่ปกติและผิดปกติออกมา

การสอน (Train) และทดสอบ (Test) ใช้ 10 - Fold Validation กล่าวคือแบ่งข้อมูลออกเป็น 10 ชุดแล้วในชุดข้อมูลแต่ละชุดแบ่งออกเป็น 10 ส่วน 9 ส่วนเพื่อสอน และอีก 1 ส่วนเพื่อทดสอบ 10 รอบ แล้วนำค่าความแม่นยำมาคำนวณได้จากอัตราส่วนระหว่างจำนวนตัวอย่างในชุดข้อมูลทดสอบที่ทำนายกลุ่มได้อย่างถูกต้องทั้งหมด ได้แผนภาพต้นไม้ดัง

ภาพที่ 18



ภาพที่ 18 ต้นไม้ตัดสินใจเพื่อแยกกลุ่มข้อมูลปกติและผิดปกติ

ต้นไม้ตัดสินใจ

ภาพที่ 18 ด้วยการเลือกคุณลักษณะโดยใช้เอนโทรปีด้วยการแต่งกิ่ง (Pruningแตกใบด้วยค่า (ชั้น 100 ค่า โดยจำกัดความลึกไม่เกิน 5 ค่า โดยหยุดแยกโหนดด้วยค่าน้อยกว่า 2 อย่างน้อย ด้วยสีเขียวคือกลุ่มข้อมูลที่ปกติ และสีแดงคือกลุ่มข้อมูลที่ผิดปกติ

5462082 18

Node 1 ซึ่งเป็นราก ประกอบด้วยข้อมูลปกติ 104,381 ระเบียน และผิดปกติ 119 ระเบียน รวมเป็น 104,ระเบียน 500

Node 2 เป็นกลุ่มข้อมูลที่ผิดปกติจำนวน 72 ระเบียน แทนด้วยสีแดงจากการตั้งเงื่อนไขการนับจำนวนคำว่า “No Data” เท่ากับสองคำ

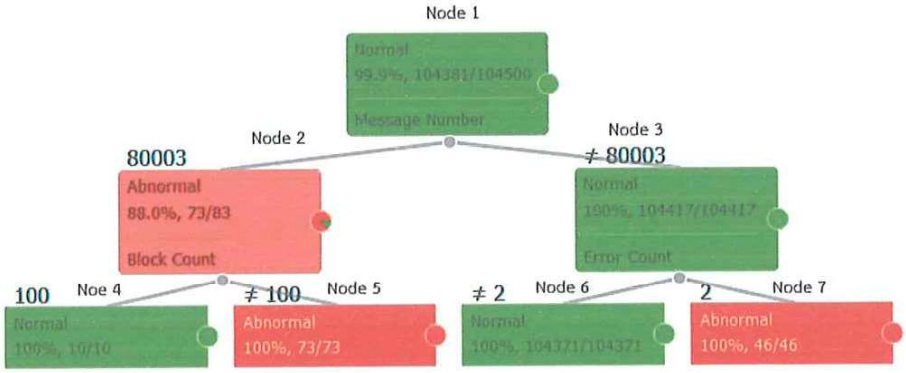
Node 3 เป็นกลุ่มข้อมูลปกติจำนวน 104,428 ระเบียน เงื่อนไขว่าไม่พบคำว่า “No Data” ในรายการการเปลี่ยนแปลง

Node 4 เป็นกลุ่มข้อมูลที่ผิดปกติจำนวน 47 ระเบียน จากเงื่อนไขจำนวนคำว่า “Error” ต่อรายการเปลี่ยนแปลงเท่ากับสองคำ

Node 5 เป็นกลุ่มข้อมูลปกติจำนวน 104,381 ระเบียน จากเงื่อนไขจำนวนคำว่า “Error” ไม่ใช่สองคำ

จากต้นไม้นี้จะพบว่าใช้เพียงสองคุณลักษณะคือการนับจำนวนคำว่า “No Data” และคำว่า “Error” เพื่อให้เข้าใจสาเหตุอย่างอื่นที่ทำให้เกิดความผิดพลาด เราจะตัดคุณลักษณะนับจำนวนคำ “No Data” ออกจากระบบเรียนรู้ ต้นไม้ตัดสินใจจะมีลักษณะเป็น

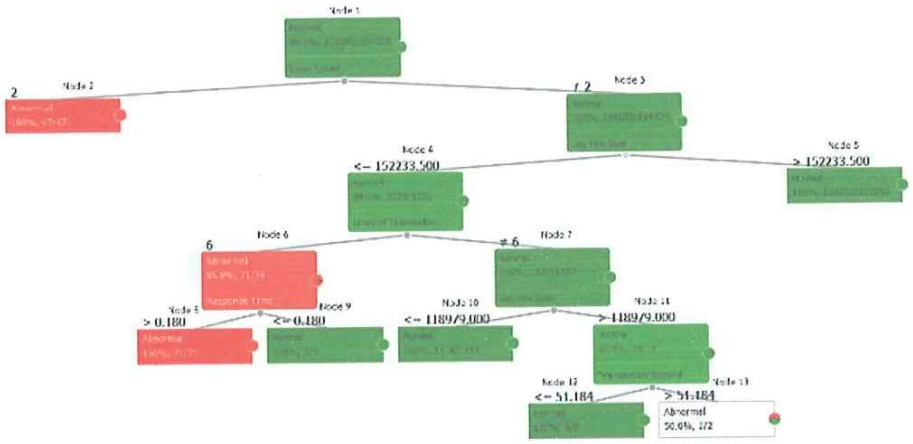
ภาพที่ 19



ภาพที่ 19 ต้นไม้สินใจที่ตัดคุณลักษณะการนับคำว่า “No Data” ออก

ต้นไม้นี้แสดงว่าคุณลักษณะแรกที่ใช้ตัดสินใจคือ หมายเลขข้อความเป็น 80003 จากนั้นเป็นการนับคำว่า “Block” และการนับคำว่า “Error” เนื่องจากหมายเลขข้อความไม่สามารถนำมาใช้ในทางปฏิบัติเนื่องจากหมายเลขจะเปลี่ยนไปตามเวลาที่เครื่องจักรทำงาน จึงตัดคุณลักษณะนี้ออก เพื่อดูให้ลึกลงไปอีกว่า มีคุณลักษณะใดอีกที่เป็นสาเหตุของความผิดพลาด แล้วได้ต้นไม้ตัดสินใจตาม

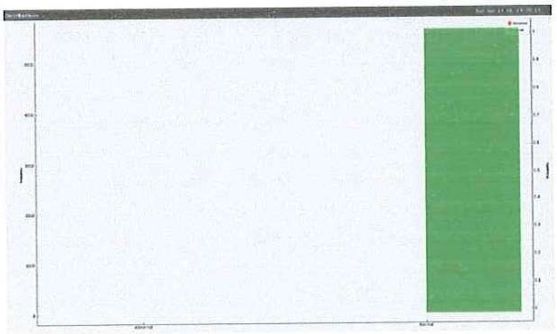
ภาพที่ 20



ภาพที่ 20 ต้นไม้สินใจที่ตัดหมายเลขข้อความออก

จะเห็นได้ว่า คุณลักษณะที่บ่งบอกความผิดพลาดคือ “error count” “log file size” “line of transaction” “response time” และ “transaction time”

หลังจากได้แบบจำลองจากต้นไม้ตัดสินใจแล้วทดลองนำแบบจำลองจากภาพที่ 21 เป็นการวิเคราะห์แฟ้มล็อกของวันที่ 2 มกราคม ค.ศ. 2016 จำนวน 5,640 ระเบียน ได้ผลถูกต้องทั้งหมดโดยใช้เวลาไม่ถึงหนึ่งนาที



ภาพที่ 21 ผลลัพธ์ของการนำแบบจำลองมาใช้จำแนกแฟ้มล็อก

5462082218

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การเพิ่มและปรับปรุงความสามารถของการประมวลผลก่อนทำให้ต้นไม้ตัดสินใจเรียนรู้ข้อมูลจริงในสายการผลิต สามารถนำมาใช้ตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกได้ถูกต้อง 100% ตามตารางที่ 6 ตารางที่ 7 และตารางที่ 8 โดยมีคุณลักษณะที่สำคัญ คือ การปรากฏคำว่า "No data" การปรากฏคำว่า "Error" ขนาดของแฟ้มล็อก จำนวนบรรทัดของเหตุการณ์ เวลาตอบสนอง และเวลาที่ใช้ในการลงบันทึก คุณลักษณะเหล่านี้สามารถอธิบายความเกี่ยวข้องกับการเกิดความผิดพลาดของเครื่องจักรในสายการผลิตได้ เช่นคำว่า "No data" สะท้อนถึงความผิดปกติที่มาจากวิศวกร และคำว่า "Error" สะท้อนถึงความผิดปกติที่มาจากระบบ ขนาดของแฟ้มล็อกที่ใหญ่สะท้อนถึงความผิดปกติของเครื่องจักร เวลาตอบสนอง และระยะเวลาที่เกิดเหตุการณ์มีส่วนบ่งบอกความผิดพลาดเช่นกัน

5462082218

การใช้โปรแกรมเพื่อตรวจสอบความผิดปกติของเครื่องจักรจากแฟ้มล็อกโดยอัตโนมัติช่วยให้การหาสาเหตุที่เครื่องจักรทำงานผิดพลาด และการแก้ไขโดยทีมวิศวกรผู้ดูแลการผลิตให้สายการผลิตดำเนินการต่อไปเป็นไป เป็นไปอย่างรวดเร็วและแม่นยำ ช่วยลดเวลาที่สายการผลิตต้องหยุดชะงักลงได้ นอกจากนี้การวิเคราะห์ต้นไม้ตัดสินใจสามารถทำให้ทีมวิศวกรเข้าใจสาเหตุที่เครื่องจักรในสายการผลิตทำงานผิดพลาดได้ดียิ่งขึ้น ซึ่งจะนำไปสู่การปรับปรุงกระบวนการผลิตให้ลดข้อผิดพลาดลงได้อีกด้วย

แนวทางในการวิจัยต่อคือการเพิ่มคุณลักษณะสำคัญอื่น เช่น ชื่อของซอฟต์แวร์ เวอร์ชันของซอฟต์แวร์ เพื่อค้นหาว่าซอฟต์แวร์แต่ละตัวมีผลกับความผิดปกติของเครื่องจักรหรือไม่ และเวลาในการติดตั้งซอฟต์แวร์ เพื่อค้นหาว่าการปรับปรุงซอฟต์แวร์มีผลกับความผิดปกติของเครื่องจักรหรือไม่ เป็นต้น

รายการอ้างอิง

[1] สำนักงานเศรษฐกิจอุตสาหกรรม, "สรุปสถานการณ์การผลิตภาคอุตสาหกรรม เดือนมีนาคม ปี 2559 และไตรมาสที่ 1 ปี 2559," *OIE SHARE,* vol. 5, pp. 6-7, พฤษภาคม 2016.

[2] สำนักวิจัยเศรษฐกิจอุตสาหกรรม, "สรุปสถานการณ์การผลิตภาคอุตสาหกรรม เดือนมกราคม ปี 2559," *OIE SHARE,* vol. 5, pp. 8-9, มีนาคม 2016.

[3] A. Dokmai and A. Kengpol, "The Improvement for Optimization of Head Stack Assembly (HSA) Assembling Process by Using the Virtual Reality 3D Simulation Model," *AIJSTPME,* vol. 3, pp. 47-56, 2010.

[4] B. Krasnove, "Western Digital Corporation - Fortune 500 1997 - Fortune," *FORTUNE,* 2014-05-15 1997.

[5] Fortune, "Western Digital Corporation - Fortune 500 2015 - Fortune," *FORTUNE,* 4 June 2015.

[6] N. Boonyakiat, "WD Profile 2007," ed, 2007.

[7] สำนักคณบดี คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, งานพัฒนาองค์กรและประสิทธิภาพ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, งานบริหารวิจัย คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, งานบริหารวิชาการ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, "แผนที่นำทางคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2030," 2014.

[8] ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (2016). โครงการ *ILP (CU-ENG_Industrial Liaison Program)* ร่วมกับ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวฯ จุฬาฯ จัดสัมมนา *ILP Forum* ครั้งที่ *1* ประจำปี *2558* เรื่อง *"Big Data" | ChulaEngineering*. Available: http://www.eng.chula.ac.th/node/2368.

[9] S. G. Eick, M. C. Nelson, and J. D. Schmidt, "Graphical analysis of computer log files," *Communications of the ACM,* vol. 37, pp. 50-56, December 1994.

[10] A. Fabio. (2016, 26 October). *Killed by a Machine: The Therac-25*. Available: http://hackaday.com/2015/10/26/killed-by-a-machine-the-therac-25/

[11] Gartner. (2016, August). *Gartner's 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor.* Available: http://www.gartner.com/newsroom/id/3114217

5462082218

[12] ญาใจ ลิ่มปิยะกรณ์, การทำเหมืองข้อมูล:ตำราวิศวฯ 100 ปี 2456 - 2556, 2559

[13] K. Tangchaichit and W. Kaewka, "A Study of the Head Stacks Assembly (HSA) during the Swaging Process: Optimization of the Ball Velocity," *KKU Research Journal,* vol. 15, pp. 104-112, February 2010.

[14] T. M. Mitchell, *Machine Learning*. Boston, New York: Boston : WCB/McGraw-Hill, 1997.

[15] W. Kaewka and K. Tangchaichit, "The characterization in the Head Stack Assembly (HSA) During the Swaging Process: Optimization of Actuator Arm Material," *KKU Research Journal,* vol. 15, pp. 910-918, October 2010.

[16] A. Siltepavet, S. Sinthupinyo, and P. Chongstitvattana, "A Heuristic Approach to Improve The Decision Tree of Product Acceptance in Hard Drive Manufacturing," presented at the The Eighth National Conference on Computing and Information Technology, Dusit Thani Hotel 240/2 Pattaya Beach Road, Pattaya City, Chonburi 20150, 2012.

[17] A. Siltepavet, S. Sinthupinyo, and P. Chongstitvatana, "Improving Quality of Products in Hard Drive Manufacturing by Decision Tree Technique," *International Journal of Computer Science Issues,* vol. 9, pp. 29-34, May 2012.

[18] Machine Learning Group at the University of Waikato. *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. Available: http://www.cs.waikato.ac.nz/ml/weka/

[19] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. NZ: University of Waikato, 2013.

[20] M. Štajdohar and J. Demšar, "Interactive Network Exploration with Orange," *Journal of Statistical Software,* vol. 53, 2013.

[21] T. Gao, C. Du, W. Sun, and L. Xie, "An integrated plant/ control design method and application in hard disk drives," *International Journal of Systems Science,* vol. 47, pp. 644-651, 23 May 2012 2016.

[22] F. Schuster, A. Paul, and H. König, "Towards Learning Normality for Anomaly Detection in Industrial Control Networks," in *AIMS*, 2013, pp. 61-72.

5462108218

[23] T. M. Mitchell, "The Discipline of Machine Learning," Carnegie Mellon University, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, Perspective July 2006.

[24] B. Schöölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High- Dimensional Distribution," *Neural Computation,* vol. 13, pp. 1443-1471, 2001.

[25] V. Roth, "Outlier detection with one-class kernel fisher discriminants," *Advances in Neural Information Processing Systems,* vol. 17, pp. 1169-1176, 2005.

[26] V. Roth, "Kernel Fisher Discriminants for Outlier Detection," *Neural Computation,* vol. 18, pp. 942-960, 2006.

[27] J. R. Quinlan, "Induction of Decision Trees," *Mach Learn,* vol. 1, pp. 81-106, 1986.

[28] J. R. Quinlan, "C4.5: programs for machine learning," ed: San Mateo, Calif.: Morgan Kaufmann Publishers, 1993.

[29] V. Lavrenko and N. Goddard, "Decision Tree," ed: School of Informatics, 2014.

[30] โอมิ เคนอิชิ, *7 จุดบอดแฝงที่ขัดขวางการเพิ่มผลผลิตของโรงงาน*, พิมพ์ครั้งที่ 1 กรุงเทพฯ : สถาบันส่งเสริมเทคโนโลยี (ไทย-ญี่ปุ่น), 2558.

[31] P. Bhoria and K. Garg, "An Imperial learning of Data Mining Classification Algorithms in Intrusion Detection Dataset Citefactor.org-Journal|Research Paper|Indexing|Impact factor," *International Journal of Scientific & Engineering Research,* vol. 4, pp. 2394-2399, 2 August 2014 2013.

[32] J. Demšar, B. Zupan, G. Leban, and T. Curk, "Orange: From Experimental Machine Learning to Interactive Data Mining," in *PKDD*, 2004, pp. 537-539.

[33] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, *et al.*, "Orange: Data Mining Toolbox in Python," *Journal of Machine Learning Research,* vol. 14, pp. 2349-2353, 2013.

[34] M. Štajdohar and J. Demšar, "Interactive Network Exploration with Orange," *Journal of Statistical Software,* vol. 53, pp. 1-24, 2013.

[35] Google Inc. (2016, 19 June). *Google Summer of Code*. Available: https://summerofcode.withgoogle.com

5462208218

[36] U. o. L. Bioinformatics Laboratory, "Orange Blog | Data Mining – Fruitful and Fun," 2016.

[37] J. H. Andrews, "Testing using log file analysis: tools, methods, and issues," ed. Honolulu, HI: IEEE, 1998, pp. 157-166.

5462082I8

ภาคผนวก

ประวัติผู้เขียนวิทยานิพนธ์

นายธนาธาร ภัทรกวิน เคยผ่านการคัดเลือกเป็นตัวแทนคอมพิวเตอร์โอลิมปิคของโรงเรียนเบ็ญจะมะมหาราชและจังหวัดอุบลราชธานีแข่งขันระดับภูมิภาคในช่วงที่ศึกษาและสำเร็จการศึกษาชั้นมัธยมศึกษา เมื่อปีการศึกษาที่ 2537 จากโรงเรียนเบ็ญจะมะมหาราช จังหวัดอุบลราชธานี และสำเร็จการศึกษาวิศวกรรมศาสตรบัณฑิต เมื่อปีการศึกษา 2543 สาขาวิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี หลังจากนั้นได้เข้าศึกษาต่อหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2557

ระหว่างศึกษาหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยได้รับทุนอุดหนุนการศึกษาจากโครงการด้านการศึกษาและโครงการให้ความช่วยเหลือของมหาวิทยาลัยสนับสนุนจาก Western Digital ภายใต้การสนับสนุนของคุณเกียรติพงษ์ หล้านัก และดร.นเรนทร์ หล่อบรรจงสุข ในความดูแลของศาสตรจารย์ ดร.ประภาส จงสถิตย์วัฒนา

5962087H