

การศึกษาเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่แน่นอนเพื่อลดเวลาในกระบวนการ
ทดสอบวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2565

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARATIVE STUDY ON UNCERTAINTY REGRESSIONMODEL FOR A WRITE CURRENT
TEST TIME REDUCTION IN HARD DRIVE MANUFACTURING PROCESSES



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics
Department of Statistics
FACULTY OF COMMERCE AND ACCOUNTANCY
Chulalongkorn University
Academic Year 2022
Copyright of Chulalongkorn University

ภัทรดิศ ดำรงค์ศักดิ์ : การศึกษาเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่
 แน่นนอนเพื่อลดเวลาในกระบวนการทดสอบวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของ
 ฮาร์ดไดรฟ์ . (A COMPARATIVE STUDY ON UNCERTAINTY REGRESSION MODEL
 FOR A WRITE CURRENT TEST TIME REDUCTION IN HARD DRIVE
 MANUFACTURING PROCESSES) อ.ที่ปรึกษาหลัก : ผศ. ดร.อนุภาพ สมบูรณ์สวัสดิ์

ฮาร์ดไดรฟ์ (HDD) เป็นอุปกรณ์บันทึกข้อมูลแม่เหล็กที่มีความแม่นยำสูง ดังนั้นจึงมี
 ค่าใช้จ่ายสูง และเสียเวลาในการวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดฮาร์ดไดรฟ์ หากจ่าย
 กระแสไฟฟ้าเขียนไม่เหมาะสมจะส่งผลกระทบต่อประสิทธิภาพการทำงานของฮาร์ดไดรฟ์ ซึ่งเราใช้
 วิธีการเงื่อนไขการทดสอบแบบปรับตัว (Adaptive Test Condition) เป็นเทคนิคที่ปรับวิธีการ
 ทดสอบแบบดั้งเดิม ตามรูปแบบข้อมูลพารามิเตอร์ เพื่อปรับปรุงวิธีการทดสอบปัจจุบัน และลด
 เวลาการทดสอบ งานวิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการใช้ตัว
 แบบจำลองการถดถอยโดยความไม่แน่นอนสำหรับการลดช่วงการวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสม
 ที่สุด สำหรับการลดเวลาการทดสอบวัดค่ากระแสไฟฟ้าเขียน (write current test) โดยการ
 คำนวณช่วงความเชื่อมั่นของผลทำนายที่ระดับความเชื่อมั่นที่ยอมรับได้ โดยใช้ค่าความไม่แน่นอน
 ของข้อมูล (Data uncertainty) ที่ผ่านวิธีปรับการเทียบมาตรฐาน (Recalibration) แล้วนำมาลด
 ช่วงวัดที่ได้จากการทดสอบฮาร์ดไดรฟ์ จากนั้นนำช่วงเชื่อมั่นของผลทำนายนั้นมาลดช่วงการวัดค่า
 กระแสไฟฟ้าเขียน โดยการศึกษา และเปรียบเทียบใช้ตัวแบบจำลองการถดถอยโดยความไม่
 แน่นนอน ได้แก่ NGBoost, XGB-Distribution และ CatBoost ซึ่งผลลัพธ์ของงานวิทยานิพนธ์
 คือ CatBoost สามารถลดเวลาในการทดสอบวัดค่ากระแสไฟฟ้าเขียนสูงสุดในช่วงความเชื่อมั่นของ
 ผลทำนาย ณ ระดับความเชื่อมั่นที่ยอมรับได้ ซึ่งครอบคลุมสัดส่วน 0.9 ของทุกชุดการทดสอบ

สาขาวิชา สถิติ
 ปีการศึกษา 2565

ลายมือชื่อนิสิต
 ลายมือชื่อ อ.ที่ปรึกษาหลัก

6480464626 : MAJOR STATISTICS

KEYWORD: Manufacturing processes, Hard drives, Uncertainty, Machine learning,
Test time reduction

Pattaradit Damrongsak : A COMPARATIVE STUDY ON UNCERTAINTY
REGRESSIONMODEL FOR A WRITE CURRENT TEST TIME REDUCTION IN
HARD DRIVE MANUFACTURING PROCESSES . Advisor: Asst. Prof. ANUPAP
SOMBOONSAVATDEE

Hard drives(HDD) are highly accurate magnetic recording devices, therefore it is costly and time-consuming to measure the optimal write current for HDD. The HDD's performance could be badly affected if the write current is invalid. The adaptive test condition is a method of adapting traditional test methods by adjusting the test in response to the pattern of parameters, to improve traditional measurement of the optimal write current test for time reduction. The purpose of this research is to study and compare different approaches to using uncertainty regression models for reducing the measurement range of the optimal write current test for test time reduction by calculating the prediction interval at an acceptable confidence level using the recalibrated data uncertainty. NGBoost, XGB-Distribution, and CatBoost are among the uncertainty regression models used. Then, each model is performed in order to determine which one maximizes the test time reduction at the accepted confidence level. The primary result, CatBoost is maximum the write current test time reduction at an acceptable level of confidence that the prediction interval covers 0.9 proportion of test sets.

Field of Study: Statistics

Student's Signature

Academic Year: 2022

Advisor's Signature

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.อนุภาพ สมบูรณ์สวัสดิ์ เป็นอย่างยิ่งที่ได้ให้โอกาสผู้วิจัยได้เป็นลูกศิษย์ในที่ปรึกษา สละเวลาให้คำแนะนำสั่งสอนและคำปรึกษาที่มีประโยชน์ อีกทั้งยังช่วยแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ ดร.วิฐรา พึ่งพาพงศ์ ประธานกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. ภูมิพันธุ์ รุจิขจร และผู้ช่วยศาสตราจารย์ ดร. ธิดาพร ศุภภากร กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาสละเวลามาตรวจทานแก้ไขข้อบกพร่องในวิทยานิพนธ์ฉบับนี้ ตลอดจนให้คำแนะนำที่เป็นประโยชน์แก่ผู้วิจัยที่ช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ผู้วิจัยขอกราบขอบพระคุณ ดร.สมชาย นวลประเสริฐ ผู้จัดการฝ่ายวิศวกรรม กลุ่มวิทยาศาสตร์ข้อมูล(DSG) และทีม กลุ่มวิทยาศาสตร์ข้อมูล(DSG) บริษัท ซีเกท เทคโนโลยี (ประเทศไทย) จำกัด โคราซ เป็นอย่างยิ่งที่ได้ให้โอกาสผู้วิจัยได้เป็นนิสิตฝึกงาน ช่วยฝึกปฏิบัติงานเพื่อพัฒนาในการเรียน และคำปรึกษาที่มีประโยชน์ อีกทั้งยังเป็นประโยชน์แก่ผู้วิจัยต่อไป

ผู้วิจัยขอกราบขอบพระคุณ บริษัท ซีเกท เทคโนโลยี (ประเทศไทย) จำกัด โคราซ ที่ให้ความอนุเคราะห์รับผู้วิจัยได้เป็นนิสิตฝึกงาน สนับสนุนอุปกรณ์สำหรับฝึกปฏิบัติงานและงานวิทยานิพนธ์ และข้อมูลสำหรับงานวิทยานิพนธ์ ตลอดจนเป็นประโยชน์แก่ผู้วิจัยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ทุกท่านที่ให้ความรู้ทางด้านวิชาการ รวมไปถึงเจ้าหน้าที่ของภาควิชาสถิติที่ช่วยจัดทำเอกสาร และอำนวยความสะดวกในด้านต่าง ๆ

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณบิดา มารดา และครอบครัวที่คอยสนับสนุนและให้กำลังใจเสมอมา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภัทรติศ ดำรงค์ศักดิ์

สารบัญ

	หน้า
.....	ก
บทคัดย่อภาษาไทย	ก
.....	ง
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
บทที่ 1 บทนำ	5
1.1 ความเป็นมาและความสำคัญของปัญหา.....	5
1.2 วัตถุประสงค์ของงานวิจัย	6
1.3 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง	8
2.1 ฮาร์ดแวร์.....	8
2.1.1 การบันทึกระบบแม่เหล็ก.....	8
2.1.2 โครงสร้างภายในของฮาร์ดแวร์	9
2.1.3 กระแสไฟฟ้าเขียน.....	10
2.1.4 การทดสอบกระแสไฟฟ้าเขียน	11
2.2 การวิเคราะห์ความถดถอย	13
2.3 การเรียนรู้ของเครื่อง (Machine Learning).....	15
2.4 Ensemble Learning.....	15
2.5 ความไม่แน่นอน(Uncertainty).....	16
2.6 ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	19

2.6.1	NGBoost (Natural Gradient Boosting).....	19
2.6.2	XGBoost-Distribution (eXtreme Gradient Boosting Distribution).....	19
2.6.3	Catboost (Gradient Boosting with categorical features support)	20
2.7	การเทียบมาตรฐาน และการปรับเทียบมาตรฐานสำหรับแบบจำลองการถดถอย	21
2.7.1	การเทียบมาตรฐานสำหรับแบบจำลองการถดถอย (Calibration Regression).....	21
2.7.2	การปรับเทียบมาตรฐานสำหรับแบบจำลองการถดถอย (Recalibration Regression)	23
2.8	วิธีการแก้ปัญหาการลดเวลาการทดสอบ	23
2.8.1	แนวคิดสำหรับแก้ปัญหาการทดสอบกระแสเขียน	23
2.8.2	เงื่อนไขการทดสอบแบบปรับตัว (Adaptive Test Condition)	25
2.9	เกณฑ์คะแนนที่ใช้ในการศึกษา และเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	26
2.9.1	MAE (Mean Absolute Error).....	26
2.9.2	RMSE (Root Mean Squared Error)	27
2.9.3	<i>AvgTTR</i>	28
บทที่ 3	วิธีการดำเนินงานวิจัย	31
3.1	ขอบเขตของงานวิจัย	31
3.2	วิธีการดำเนินการวิจัย.....	31
3.2.1	นำเข้าข้อมูล และทำความสะอาดข้อมูลสำหรับเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน.....	31
3.2.2	สำรวจ และวิเคราะห์ข้อมูลเชิงลึกของชุดข้อมูล	34
3.2.3	เตรียมการชุดข้อมูลให้พร้อมสำหรับเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	35
3.2.4	ดำเนินฝึกการเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	37

บทที่ 4 ผลการวิจัย.....	38
ข้อกำหนดเบื้องต้น.....	38
ข้อจำกัดของงานวิจัย	39
4.1 ผลลัพธ์การประมาณแบบจำลองการถดถอยโดยความไม่แน่นอน ก่อนนำค่าความไม่แน่นอน ของข้อมูลไปใช้คำนวณ <i>AvgTTR</i>	40
4.2 ผลลัพธ์การเปรียบเทียบความผิดพลาดของการเทียบมาตรฐาน (Miscalibration Area).....	41
4.3 ผลการเปรียบเทียบความสามารถในการลดเวลาการทดสอบกระแสไฟฟ้าเขียนของชุดทดสอบ แต่ระดับความเชื่อมั่น.....	46
4.4 ผลการเปรียบเทียบความสามารถในการลดเวลาการทดสอบกระแสไฟฟ้าเขียนของชุดทดสอบ และชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ.....	48
บทที่ 5 สรุปผลการดำเนินงาน และข้อเสนอแนะ	54
5.1 สรุปและอภิปรายผลการดำเนินงาน	54
5.2 ข้อเสนอแนะ	55
บรรณานุกรม	57
ประวัติผู้เขียน	60

สารบัญรูปภาพ

หน้า

รูปที่ 1 ภาพการใช้วิธี Adaptive Test Condition ในการลดเวลาในการทดสอบวัดค่ากระแสไฟฟ้า เขียน คัดแปลงจากการประมวลสัญญาณสำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 :พื้นฐานของสัญญาณ อ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์ แห่งชาติ	6
รูปที่ 2 โครงสร้างภายในของฮาร์ดไดรฟ์ จากหนังสือการประมวลสัญญาณสำหรับการจัดเก็บข้อมูล ดิจิทัล เล่ม 1 :พื้นฐานของสัญญาณอ่าน (หน้า 7), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยี อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ	9
รูปที่ 3 โครงสร้างภายในของฮาร์ดไดรฟ์ จากหนังสือการประมวลสัญญาณสำหรับการจัดเก็บข้อมูล ดิจิทัล เล่ม 1 :พื้นฐานของสัญญาณอ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยี อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ	9
รูปที่ 4 สัญญาณกระแสเขียนที่วงจร preamplifier จาก ผลกระทบของกระแสเขียนต่อสมรรถนะของ ระบบบันทึกข้อมูลเชิงแม่เหล็กแนวตั้ง ,เรื่องเต็มการประชุมทางวิชาการของ มหาวิทยาลัยเกษตรศาสตร์ ครั้งที่48 ,โดย พรชัย เปลี่ยมทรัพย์ และคณะ ,2553 ,กรุงเทพฯ ,หน้า 302-309 (527 หน้า).....	10
รูปที่ 5 กระแสไฟฟ้าเขียนที่จ่ายสู่หัวเขียน	11
รูปที่ 6 ภาพการกระบวนกรทดสอบวัดค่ากระแสไฟฟ้าเขียน คัดแปลงจากการประมวลสัญญาณ สำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 :พื้นฐานของสัญญาณอ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ	12
รูปที่ 7 ลักษณะการทำงานของ Boosting	15
รูปที่ 8 การเลือกชุดทดสอบที่เหมาะสมที่สุด(Optimal test set).....	25
รูปที่ 9 การตรวจจับตัวอย่างที่มีข้อบกพร่อง(Detecting defective samples).....	25
รูปที่ 10 การลดช่วงค้นหาในการทดสอบของพารามิเตอร์(Reducing search space of parameter).....	26
รูปที่ 11 กราฟแท่งที่แสดงอัตราส่วนของจำนวนตัวอย่างของแต่ละช่วงความกว้าง.....	34

รูปที่ 12 การแจกแจงของตัวแปรตาม , รูปที่ 13 การแจกแจงของตัวแปรตามแต่ละช่วงความกว้างวัด ค่า	34
รูปที่ 14 ความสัมพันธ์ระหว่างตัวแปรอิสระโดยวิธี Pearson	35
รูปที่ 15 แผนผังการวิธีการวิจัย	37
รูปที่ 16 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบ	41
รูปที่ 17 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ	42
รูปที่ 18 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบ	42
รูปที่ 19 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration	44
รูปที่ 20 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration	44
รูปที่ 21 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration	45
รูปที่ 22 แสดงการลดเวลาการทดสอบแต่ละตัวแบบจำลอง แต่ละระดับความเชื่อมั่น	46
รูปที่ 23 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration	48
รูปที่ 24 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration	48
รูปที่ 25 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่าง ผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration	49

รูปที่ 26 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration	49
รูปที่ 27 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration	50
รูปที่ 28 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration	50
รูปที่ 29 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration	51
รูปที่ 30 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration	51
รูปที่ 31 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration	52

สารบัญตาราง

หน้า

ตารางที่ 1 แสดงชื่อตัวแปร ความหมายของตัวแปร และประเภทตัวอย่างของตัวแปร แต่ละตัวอย่าง	32
ตารางที่ 2 จำนวนหน่วยตัวอย่างแต่ละสัปดาห์	36
ตารางที่ 3 ผลลัพธ์เกณฑ์คะแนน MAE และ RSME ของชุดฝึก และทุกชุดตัวอย่างทดสอบ แต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	40
ตารางที่ 4 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ของตัวแบบจำลองความไม่แน่นอน ก่อน Recalibration	43
ตารางที่ 5 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ของตัวแบบจำลองความไม่แน่นอน หลัง Recalibration	45
ตารางที่ 6 ผลลัพธ์การลดเวลาการทดสอบ แต่ละระดับความเชื่อมั่น	47
ตารางที่ 7 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ และการลดเวลาการทดสอบของตัวแบบจำลองความไม่แน่นอน แต่ละชุดตัวอย่างทดสอบ	52
ตารางที่ 8 ผลลัพธ์เกณฑ์คะแนน Miscalibration Area ก่อนทำ และหลังทำ Recalibration ในบทที่ 4.2	54
ตารางที่ 9 ผลลัพธ์เกณฑ์การลดเวลาการทดสอบของตัวแบบจำลองความไม่แน่นอน แต่ละชุดตัวอย่างทดสอบ ในบทที่ 4.4	55

บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ฮาร์ดไดรฟ์ (HDD) คือ อุปกรณ์จัดเก็บข้อมูลดิจิทัลแบบไม่ลบเลือนที่จัดเก็บข้อมูลบนดิสก์แม่เหล็กที่หมุนได้ แล้วฮาร์ดไดรฟ์นั้นยังคงเป็นอุปกรณ์การบันทึกในระบบแม่เหล็กหลักในการจัดเก็บข้อมูลจำนวนมากเนื่องจากมีประสิทธิภาพสูง ความน่าเชื่อถือสูง และคุ้มค่า ซึ่งเป็นส่วนหนึ่งของสังคมยุคใหม่ [1]

ในกระบวนการผลิตฮาร์ดไดรฟ์โดยทั่วไป หลังประกอบฮาร์ดไดรฟ์เรียบร้อยแล้วจะถูกนำไปสู่กระบวนการทดสอบประสิทธิภาพการทำงาน (Drive Test Process) จากนั้นเพื่อวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดสำหรับฮาร์ดไดรฟ์ จากช่วงการวัดค่าที่ได้จากกระบวนการทดสอบประสิทธิภาพการทำงาน ซึ่งเราเรียกวินี้ว่า การทดสอบวัดค่ากระแสไฟฟ้าเขียน (Write Current Test) ซึ่งมีค่าใช้จ่ายสูง และเสียเวลา เนื่องจากฮาร์ดไดรฟ์เป็นอุปกรณ์การบันทึกในระบบแม่เหล็กแม่นยำสูง [1] [2] หากจ่ายกระแสไฟฟ้าเขียนไม่เหมาะสมจะส่งผลกระทบต่อประสิทธิภาพการทำงานของฮาร์ดไดรฟ์ และช่วงในการวัดเดิมนั้นมีระยะที่กว้าง จึงต้องการค้นหาวินิที่จะสามารถลดเวลาในการทดสอบวัดค่ากระแสไฟฟ้าเขียน โดยเรียกวินิดังเดิมนี้ว่า “Non-Adaptive Test”

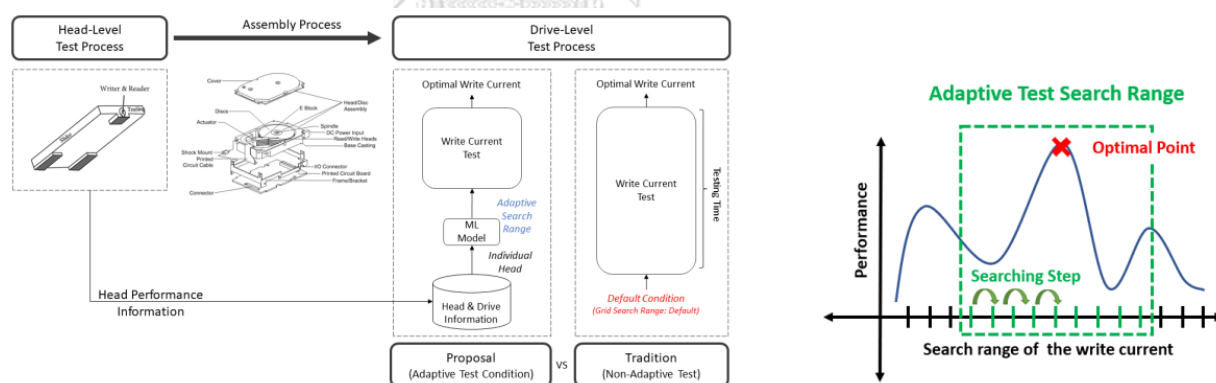
ดังนั้นการใช้เทคนิคการประมาณค่าแบบจุดของตัวแบบจำลองการถดถอย จึงไม่เหมาะสมเพราะความเสี่ยงที่ตัวประมาณจุดอาจไม่ใช่ค่ากระแสการไฟฟ้าเขียนที่เหมาะสมที่สุด เงื่อนไขการทดสอบแบบปรับตัว (Adaptive Test Condition) เป็นเทคนิคที่ปรับวิธีการทดสอบแบบดั้งเดิม ตามรูปแบบข้อมูลพารามิเตอร์ เพื่อปรับปรุงวิธีการทดสอบปัจจุบันและลดเวลาการทดสอบ [3] โดยทั่วไปแล้ววิธีของเงื่อนไขการทดสอบแบบปรับตัวมีอยู่ด้วยกัน 3 วิธี 1.การเลือกชุดทดสอบที่เหมาะสมที่สุด (Optimal test set) 2.การตรวจจับตัวอย่างที่มีข้อบกพร่อง (Detecting defective samples) 3. การลดช่วงค้นหาในการทดสอบของพารามิเตอร์ (Reducing search space of parameter)

งานวิจัยนี้เสนอการใช้ Adaptive Test Condition โดยใช้แนวทางที่ 3 เพื่อลดพื้นที่การค้นหาสำหรับการลดเวลาทดสอบในการผลิตฮาร์ดไดรฟ์ โดยนำช่วงความเชื่อมั่นของผลทำนายที่ได้มาจากความไม่แน่นอนของข้อมูล (Data Uncertainty) ที่ผ่านวิธีปรับการเทียบมาตรฐาน (Recalibration) แล้วนำมาลดช่วงวัดที่ได้จากการทดสอบฮาร์ดไดรฟ์ โดยจะปรับขอบล่างของช่วงจากการจ่ายกระแสไฟฟ้าเขียนเดิมมาเป็นขอบล่างของช่วงความเชื่อมั่นของผลทำนาย ถ้าขอบล่างของช่วงความเชื่อมั่นของผลทำนายมีค่ามากกว่าขอบล่างของช่วงจากการจ่ายกระแสไฟฟ้าเขียนเดิม และจะปรับขอบบนของช่วงจากการจ่ายกระแสไฟฟ้าเขียนเดิมมาเป็นขอบบนของช่วงความเชื่อมั่นของผล

ทำนาย ถ้าขอบบนของช่วงความเชื่อมั่นของผลทำนายมีค่าน้อยกว่าขอบบนของช่วงจากการจ่ายกระแสไฟฟ้าเขียนเดิม หากช่วงความกว้างในการวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดลดลงจากเดิม แล้วเวลาที่ใช้ในการทดสอบวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดจะลดลงด้วย

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อศึกษา และเปรียบเทียบวิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอน เพื่อลดเวลาในการทดสอบวัดค่ากระแสไฟฟ้าเขียนโดยรูปที่ 1 ในงานวิทยานิพนธ์นี้สนใจใช้วิธีเงื่อนไขการทดสอบแบบปรับตัว(Adaptive Test Condition) เพื่อลดช่วงในการวัดกระแสไฟฟ้าเขียนลง ดังนั้นต้องการตัวแบบจำลองที่สามารถให้ค่าของประมาณที่สามารถนำไปใช้เพื่อลดช่วงในการวัดที่ได้ ค่าที่เราสนใจที่จะนำมาใช้คือ ค่าความไม่แน่นอนของข้อมูล ที่ได้มาจากการเรียนรู้ของวิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอน โดยจะนำค่าความไม่แน่นอนของข้อมูลที่ได้จากเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน นำไปปรับการเทียบมาตรฐานแล้วนำมาคำนวณช่วงความเชื่อมั่นของผลทำนายที่ระดับช่วงเชื่อมั่นที่ยอมรับได้ จากนั้นนำช่วงเชื่อมั่นของผลทำนายนั้นมาลดช่วงการวัดค่ากระแสไฟฟ้าเขียน โดย ตัวแบบจำลองการถดถอยโดยความไม่แน่นอนที่ถูกเลือกในการศึกษา และเปรียบเทียบ ได้แก่ NGBoost, XGBoost-Distribution, CatBoost



รูปที่ 1 ภาพการใช้วิธี Adaptive Test Condition ในการลดเวลาในการทดสอบวัดค่ากระแสไฟฟ้าเขียน ดัดแปลงจากการประมวลสัญญาณสำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 :พื้นฐานของสัญญาณอ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1. วิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอนสามารถลดระยะเวลาในการทดสอบวัดค่ากระแสไฟฟ้าเขียนลงได้
2. เป็นแนวทางสำหรับใช้วิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอนสามารถลดระยะเวลาในการวัดค่า และทดสอบประเภทอิเล็กทรอนิกส์ไดรฟ์
3. เป็นแนวทางสำหรับการแก้โจทย์ปัญหาในลักษณะตัวแปรตามเป็นช่วงจำนวนที่มีขอบบนและขอบล่างที่อยู่ในจำนวนจริง



บทที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

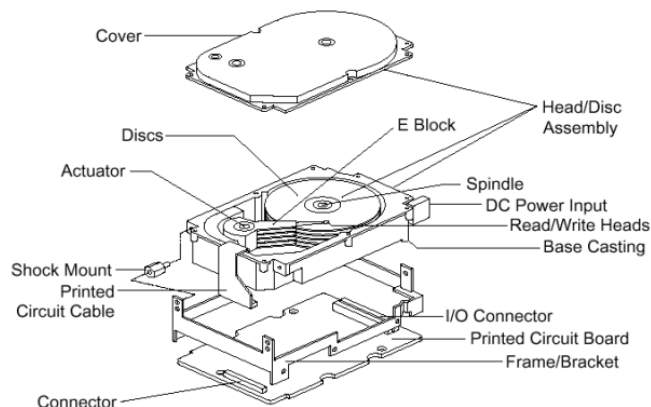
2.1 ฮาร์ดแวร์

คืออุปกรณ์จัดเก็บข้อมูลดิจิทัลแบบไม่ลบเลือนที่จัดเก็บข้อมูลบนดิสก์แม่เหล็กที่หมุนได้ แล้วฮาร์ดแวร์นั้นยังคงเป็นอุปกรณ์การบันทึกที่ระบบแม่เหล็กหลักในการจัดเก็บข้อมูลจำนวนมาก เนื่องจากมีประสิทธิภาพสูง ความน่าเชื่อถือสูง และคุ้มค่า ซึ่งเป็นส่วนหนึ่งของสังคมยุคใหม่ [1]

2.1.1 การบันทึกในระบบแม่เหล็ก

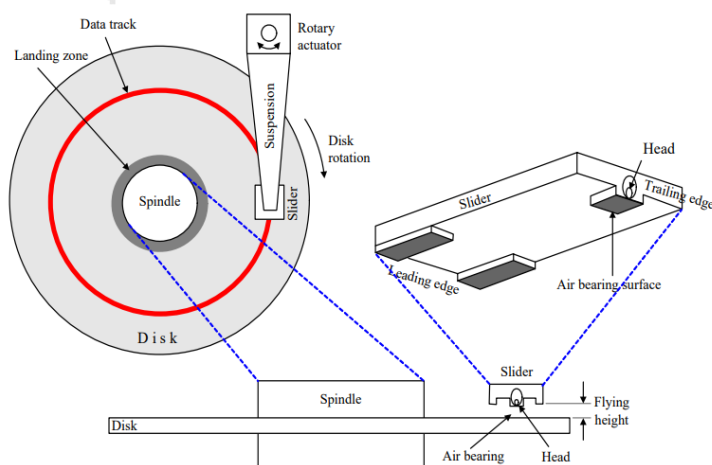
การบันทึกในระบบแม่เหล็ก คือการจัดเก็บข้อมูลบิตให้อยู่ในรูปของการเปลี่ยนแปลงระดับสภาพความเป็นแม่เหล็กในสื่อบันทึก ซึ่งสามารถแบ่งออกเป็น 2 แบบ คือ แบบแอนะล็อก และแบบดิจิทัล การบันทึกในระบบแม่เหล็กแบบแอนะล็อกจะเก็บสัญญาณ ในรูปของการเปลี่ยนแปลงระดับสภาพความเป็นแม่เหล็กอย่างต่อเนื่อง โดยระดับสภาพความเป็นแม่เหล็กจะเป็นสัดส่วนกับระดับของสัญญาณที่ถูกจัดเก็บ ตัวอย่างเช่น การเก็บสัญญาณเสียงจากไมโครโฟน เป็นต้น ในทางตรงกันข้าม การบันทึกในระบบแม่เหล็กแบบดิจิทัลจะใช้ประโยชน์จากคุณสมบัติของความเป็นแม่เหล็กของวัสดุบางชนิด ที่เมื่ออยู่ในสถานะอิ่มตัวแล้ว จะทำให้สภาพความเป็นแม่เหล็กมีทิศทางชี้ไปในทิศทางใดทิศทางหนึ่ง หรือในทิศทางตรงกันข้าม ซึ่งลักษณะการบันทึกข้อมูลแบบนี้เหมาะสำหรับการเก็บข้อมูลดิจิทัลที่มีสองสถานะ คือ บิต “1” และบิต “0” หรือเรียกกันว่า “ข้อมูลไบนารี (Binary data)” ดังนั้น วัสดุเหล่านี้จึงถูกนำมาทำเป็นสื่อบันทึกเพื่อเก็บข้อมูล เนื่องจากข้อมูลในปัจจุบันส่วนมากจะอยู่ในรูปของข้อมูลดิจิทัล เช่น ข้อมูลในเครื่องคอมพิวเตอร์ และข้อมูลที่รับส่งผ่านเครือข่ายอินเทอร์เน็ต เป็นต้น นอกจากนี้ข้อมูลแอนะล็อกก็สามารถที่จะถูกแปลงให้เป็นข้อมูลดิจิทัลได้ เพื่อให้อยู่ในรูปที่ง่ายต่อการจัดเก็บข้อมูล โดยผ่านขั้นตอนการกล้ำรหัสพัลส์ (PCM: Pulse Code Modulation) เพราะฉะนั้น การบันทึกในระบบแม่เหล็กแบบดิจิทัลจึงเหมาะสมกับการเก็บข้อมูลในปัจจุบัน [1]

2.1.2 โครงสร้างภายในของฮาร์ดไดรฟ์



รูปที่ 2 โครงสร้างภายในของฮาร์ดไดรฟ์ จากหนังสือการประมวลสัญญาณสำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 :พื้นฐานช่องสัญญาณอ่าน (หน้า 7), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

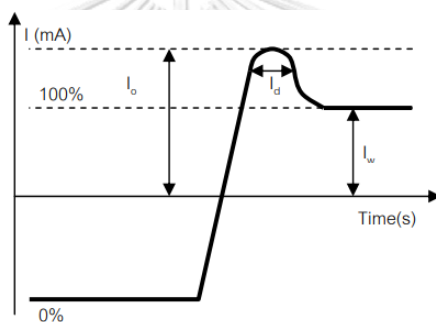
จากรูปที่ 2 แสดงให้เห็นถึงภาพโครงสร้างภายในของฮาร์ดไดรฟ์ ซึ่งมีหลักการทำงานโดยทั่วไปดังนี้ สัญญาณจะผ่านเข้าออกทางหัวอ่าน และหัวเขียน (Read/Write heads) โดยชิ้นส่วนอุปกรณ์อิเล็กทรอนิกส์ต่างๆ จะอยู่บนแผงควบคุมวงจรไฟฟ้า (Printed circuit board) ที่ติดอยู่กับฝาครอบด้านล่างของฮาร์ดไดรฟ์ โดยที่ตัวไดรฟ์ (Drive) จะถูกปิดผนึกอย่างดีเพื่อป้องกันสิ่งเปราะเปื้อน (Contamination) จากภายนอกที่จะเข้ามากระทบกับอุปกรณ์ภายในฮาร์ดไดรฟ์ อุปกรณ์อิเล็กทรอนิกส์เหล่านี้จะทำหน้าที่ควบคุมสัญญาณที่รับ และส่งจากหัวแม่เหล็ก, มอเตอร์สปินเดิล (Spindle motor) และตัวควบคุมการเคลื่อนไหว (Actuator)



รูปที่ 3 โครงสร้างภายในของฮาร์ดไดรฟ์ จากหนังสือการประมวลสัญญาณสำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 :พื้นฐานช่องสัญญาณอ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

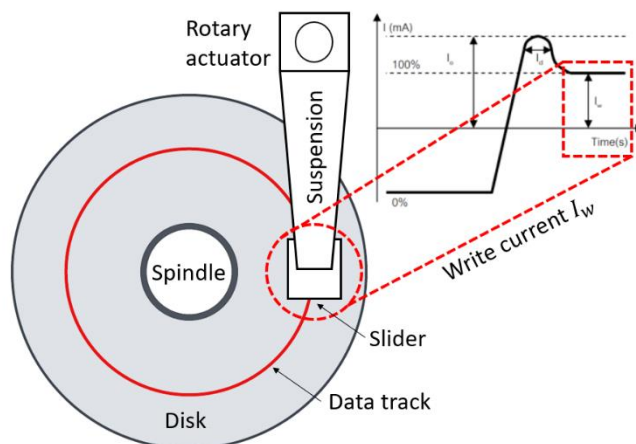
ตัวควบคุมการเคลื่อนไหวยังจะทำหน้าที่ควบคุมการเคลื่อนที่ของปีกนก (Suspension) รูปที่ 3 โดยตัวควบคุมการเคลื่อนไหวยังจะทำหน้าที่ไปรอบจุดหมุนที่กำหนด เพื่อที่จะควบคุมการวางตำแหน่งของหัวแม่เหล็กไปยังบริเวณต่างๆ ในแนวรัศมีที่เรียกว่า “โซน (Zone)” บนจานบันทึกที่เรียกว่า “ดิสก์ (Disk)” ซึ่งหัวแม่เหล็กนี้จะบินอยู่เหนือพื้นผิวของแผ่นดิสก์เป็นระยะทางที่น้อยมาก (น้อยกว่า 10^{-6} นิ้ว) ในขณะที่มอเตอร์สปินเดิลจะหมุนแผ่นดิสก์ด้วยความเร็วที่คงที่ในระหว่างที่ตัวไดรฟ์ทำงานอยู่ เมื่อตัวไดรฟ์หยุดทำงาน หรือเข้าสู่ภาวะหลับ (Sleep mode) สปินเดิลจะหยุดหมุน และตัวควบคุมการเคลื่อนไหวยังจะทำการเคลื่อนที่หัวแม่เหล็กไปยังบริเวณโซนที่ปักหัวอ่าน (Landing zone) ที่อยู่บริเวณเส้นผ่านศูนย์กลางในแผ่นดิสก์ หรือที่อยู่บนทางลาดบริเวณเส้นผ่านศูนย์กลางภายนอกของแผ่นดิสก์ เพื่อที่จะยกหัวแม่เหล็กให้อยู่ห่างจากพื้นผิวของแผ่นดิสก์ [1] [4] [5]

2.1.3 กระแสไฟฟ้าเขียน



รูปที่ 4 สัญญาณกระแสเขียนที่วงจร preamplifier จาก ผลกระทบของกระแสเขียนต่อสมรรถนะของระบบบันทึกข้อมูลเชิงแม่เหล็ก แนวตั้ง ,เรื่องเต็มการประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่48 ,โดย พรชัย เปลี่ยมทรัพย์ และคณะ ,2553 , กรุงเทพฯ ,หน้า 302-309 (527 หน้า)

รูปที่ 4 สัญญาณกระแสเขียนที่วงจร Preamplifier ป้อนให้กับหัวเขียนโดยที่แกน นอนคือ แกนของเวลาและแกนตั้งคือแกนของกระแส วงจร Preamplifier จะแปลงบิตข้อมูลให้อยู่ในรูปของกระแส เขียนและส่งผ่านไปยังขดลวดของหัวเขียน ทำให้เกิดการเหนี่ยวนำเป็นสนามแม่เหล็กเขียน บริเวณช่องว่าง (Gap) ระหว่างสื่อบันทึกกับหัวบันทึกในการบันทึกข้อมูล สัญญาณกระแสเขียนจะมี พารามิเตอร์ที่เกี่ยวข้อง 3 ตัวแปรได้แก่ แอมพลิจูดของกระแส (Write current Amplitude: I_w) ส่วน ฟุงเกินกระแส (Current Overshoot: I_o) และ เวลาของส่วนฟุงเกินกระแส (Current Duration: I_d)



รูปที่ 5 กระแสไฟฟ้าเขียนที่จ่ายสู่หัวเขียน

จากรูปที่ 5 แอมพลิจูดของกระแส คือค่าที่เราสนใจในวิทยานิพนธ์นี้จะเรียกว่า “กระแสไฟฟ้าเขียน” คือกระแสที่จ่ายไปเพื่อลบล้างสภาพความเป็นแม่เหล็กบนสื่อบันทึกหรือทำให้มีการเปลี่ยนขั้วของบิตข้อมูลตาม I_w ที่จ่ายให้กับหัวเขียน I_w ที่ถูกป้อนไปยังขดลวดของหัวเขียนจะทำให้เกิดสนามแม่เหล็กเขียนบริเวณช่องว่างระหว่างสื่อบันทึกกับหัวเขียน และสนามแม่เหล็กเขียนจะต้องมีขนาดหรือความเข้มมากกว่าสภาพลบล้างความเป็นแม่เหล็กของสื่อบันทึก ถ้าค่า I_w ต่ำกว่าจะทำให้สภาพความเป็นแม่เหล็กยังคง หลงเหลืออยู่ เป็นผลทำให้ทิศทางของเกรน (grain) ในสื่อบันทึกไม่เป็นไปตามที่ควรจะเป็น และเมื่อค่า I_w สูงขึ้นทำให้เกรนในสื่อบันทึกมีทิศทางที่ควรจะเป็น แต่ก็ทำให้หัวบันทึกยึดออกเข้าไปใกล้กับสื่อบันทึกมากขึ้นด้วย เป็นผลทำให้สนามแม่เหล็กเขียนที่ถูกสร้างขึ้นจากหัวเขียนนั้นไปกระทบกับแตร็กข้างเคียง เมื่อหัวอ่าน อ่านสัญญาณ จะได้รับสัญญาณจากแตร็กข้างเคียงกลับมาด้วยหรือที่เรียกว่า Adjacent Track Interference [1] [2]

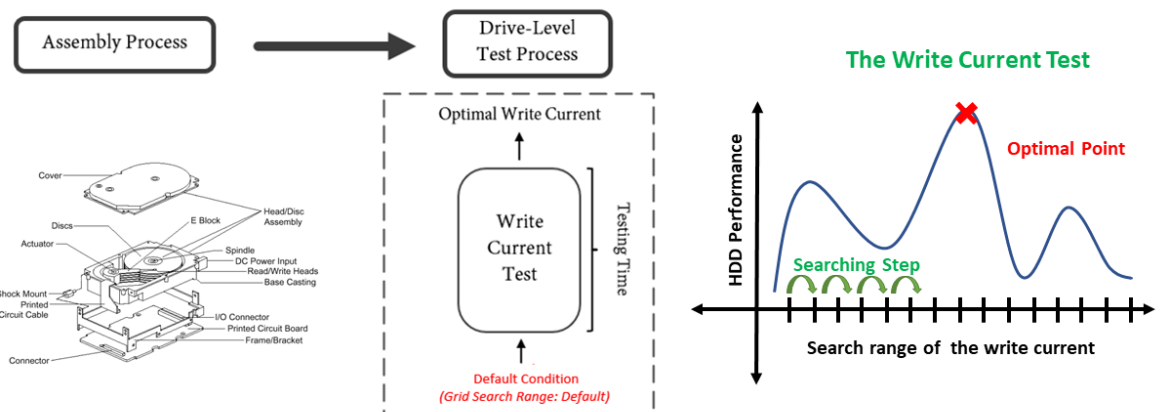
2.1.4 การทดสอบกระแสไฟฟ้าเขียน

สมรรถนะของระบบบันทึกข้อมูลเชิงแม่เหล็กแนวตั้ง เป็นแปรสำคัญในการออกแบบ และการหาค่าเหมาะสมที่สุดของผลิตภัณฑ์ฮาร์ดดิสก์ไดรฟ์ อัตราบิตผิดพลาด (Bit Error Rate) หรือ BER จะเป็นตัวชี้วัดคุณภาพ และความน่าเชื่อถือของผลิตภัณฑ์ฮาร์ดดิสก์ไดรฟ์ การออกแบบ และการออกพติไม่ซ์ผลิตภัณฑ์จะขึ้นอยู่กับหลายส่วนด้วยกัน เช่น สื่อบันทึก (Media) หัวอ่าน-เขียน (Head) เป็นต้น

ความหนาแน่นเชิงพื้นที่ (Areal density) บนสื่อบันทึก (Media) เป็นปัจจัยสำคัญในการออกแบบการเพิ่ม ความจุ (Capacity) ของฮาร์ดไดรฟ์ ซึ่งมีหลายองค์ประกอบที่เกี่ยวข้องกัน เช่น ความหนาแน่นเชิงเส้น (Linear density) ความหนาแน่นเชิงแตร็ก (Track density) หัวเขียน (Write

head) และกระแสเขียน เป็นต้น ซึ่งในการหาค่าเหมาะสมที่สุดของความหนาแน่นเชิงพื้นที่ที่จะใช้ความหนาแน่นเชิงเส้น ความหนาแน่นเชิงแตร็ก และกระแสเขียน เป็นตัวปรับให้ได้ความหนาแน่นเชิงพื้นที่ที่เหมาะสม [1]

การทดสอบวัดค่ากระแสการไฟฟ้าเขียน หรือเราเรียกว่า “Non-Adaptive Test” จากรูป 6 การทดสอบเพื่อกำหนดค่ากระแสการไฟฟ้าเขียนที่เหมาะสมที่สุด ที่ช่วยให้หัวอ่าน และหัวเขียน เขียนข้อมูลในความหนาแน่นของพื้นที่สูงได้ ฮาร์ดไดรฟ์ได้รับการทดสอบด้วยค่ากระแสไฟฟ้าเขียนแบบต่างๆ และผลกระทบของแต่ละค่าต่อพารามิเตอร์ เช่น ความหนาแน่นของแตร็ก ความหนาแน่นเชิงเส้น และวัดสัญญาณรบกวนที่อยู่ติดกัน ซึ่งกระแสไฟฟ้าเขียนที่เหมาะสมที่สุดค่าถูกเลือกสำหรับใช้ในฮาร์ดไดรฟ์ขึ้นอยู่กับพารามิเตอร์เหล่านี้ ดังนั้นระหว่างการทดสอบวัดค่ากระแสไฟฟ้าเขียนในปัจจุบันจึงใช้เวลานาน เนื่องจากมีค่ากระแสไฟฟ้าเขียนวัดค่าหลายค่า [2]



รูปที่ 6 ภาพการกระบวนการทดสอบวัดค่ากระแสไฟฟ้าเขียน ดัดแปลงจากการประมวลสัญญาณสำหรับการจัดเก็บข้อมูลดิจิทัล เล่ม 1 : พื้นฐานช่องสัญญาณอ่าน (หน้า 8), โดย รศ.ดร.ปิยะ โควินท์ทวีวัฒน์, 2550, เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

2.2 การวิเคราะห์ความถดถอย

ให้	y'_i	คือ ค่าของตัวแปรตาม ลำดับที่ i
	y_i	คือ ค่าสังเกตของตามแปรตาม ลำดับที่ i
	\hat{y}_i	คือ ค่าประมาณของตัวแปรตาม ลำดับที่ i
	$x_{j,i}$	คือ ค่าตัวแปรอิสระตัวที่ j ลำดับที่ i
	β_j	คือ ค่าพารามิเตอร์ของการวิเคราะห์ความถดถอยพหุเชิงเส้นตัวที่ j
	$\hat{\beta}_j$	คือ ค่าประมาณของค่าพารามิเตอร์ของการวิเคราะห์ความถดถอยพหุเชิงเส้นตัวที่ j
	β_0	คือ ค่าระยะตัดแกน y ที่แท้จริง
	$\hat{\beta}_0$	คือ ค่าประมาณของระยะตัดแกน y
	ε_i	คือ ค่าความคลาดเคลื่อนที่แท้จริงลำดับที่ i ซึ่งมีค่าเท่ากับ $y_i - y'_i$
	e_i	คือ ค่าประมาณความคลาดเคลื่อนลำดับที่ i ซึ่งมีค่าเท่ากับ $y_i - \hat{y}_i$
	p	คือ จำนวนตัวแปรอิสระ
	n	คือ จำนวนตัวอย่าง
	i	คือ ดัชนีบอกลำดับหน่วยตัวอย่าง ซึ่ง $i = 1, 2, \dots, n$
	j	คือ ดัชนีบอกลำดับตัวแปรอิสระ ซึ่ง $j = 1, 2, \dots, p$

การวิเคราะห์การถดถอย คือ การศึกษาความสัมพันธ์ระหว่างตัวแปรตาม กับตัวแปรอิสระ โดยที่ตัวแปรตามเป็นข้อมูลเชิงปริมาณ ส่วนตัวแปรตามเป็นได้ทั้งข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพ และสร้างสมการพยากรณ์ในการประมาณค่าเฉลี่ยของตัวแปรตามเมื่อทราบค่าของตัวแปรอิสระ โดยมีตัวแบบการวิเคราะห์ดังนี้

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_p x_{p,i} + \varepsilon_i$$

แสดงในรูปเมทริกซ์ได้ดังนี้

$$\mathbf{Y}' = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix}_{n \times 1}, \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, \hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n \times 1}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1) \times 1}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}_{(p+1) \times 1}, \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{p,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{p,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \cdots & x_{p,n} \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n \times 1}$$

ค่าสังเกตของตัวแบบของการวิเคราะห์ความถดถอยพหุเชิงเส้นสามารถเขียนได้ 2 แบบ ดังนี้

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_p x_{p,i} + \varepsilon_i$$

เขียนในรูปเมทริกซ์เป็น $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\text{หรือ } y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_p x_{p,i} + e_i$$

เขียนในรูปเมทริกซ์เป็น $\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$

และสมการที่ใช้พยากรณ์ซึ่งสร้างจากการวิเคราะห์ความถดถอยพหุเชิงเส้น คือ

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_p x_{p,i}$$

เขียนในรูปเมทริกซ์เป็น $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

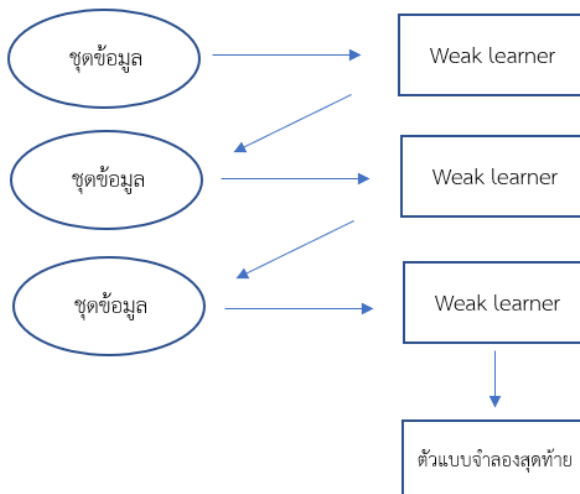
คือ รูปแบบหนึ่งของการวิเคราะห์ข้อมูล ซึ่งเป็นสาขาหนึ่งของปัญญาประดิษฐ์ เกี่ยวข้องกับการศึกษา และสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายผลข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยแบบจำลองที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์

โดยในงานวิทยานิพนธ์ นี้จะการเรียนรู้ของเครื่องประเภท การเรียนรู้แบบมีผู้สอน (Supervised Learning) กล่าวคือ ข้อมูลตัวอย่างของตัวแปรอิสระ และตัวแปรตาม เรียกว่า "ผู้สอน" ต้องถูกป้อนเข้าสู่คอมพิวเตอร์ เพื่อสร้างกฎทั่วไปที่สามารถเชื่อมโยงตัวแปรอิสระ กับตัวแปรตามได้ [6]

2.4 Ensemble Learning

เป็นการนำตัวแบบจำลองมาเรียนรู้หลาย ๆ รอบ เพื่อเพิ่มประสิทธิภาพของตัวแบบจำลอง โดยทั่วไปวิธีที่นิยมใช้ได้แก่ Bagging และ Boosting ในงานวิทยานิพนธ์ นี้จะใช้วิธีของ Boosting

ซึ่ง Boosting เป็นวิธีการเพื่อลดข้อผิดพลาดในการทำนาย โดยนำ Weak learner ที่มีความแม่นยำต่ำมาทำนายชุดข้อมูล จากนั้นให้ Weak learner ตัวใหม่มาแก้ไขค่าคลาดเคลื่อนที่มีอยู่ของ Weak learner ตัวก่อนหน้า จะดำเนินการต่อไปเรื่อย ๆ จนได้ตัวแบบจำลองที่ค่าคลาดเคลื่อนรวมที่น้อยที่สุด [7] ลักษณะการทำงานของ Boosting ดังรูปที่ 7



รูปที่ 7 ลักษณะการทำงานของ Boosting

2.5 ความไม่แน่นอน(Uncertainty)

เราจะพิจารณาค่าความไม่แน่นอนโดย Uncertainty Estimation via Bayesian Ensembles อ้างจาก Standard Bayesian Ensemble-based framework โดยกำหนดให้ θ เป็นพารามิเตอร์ของตัวแบบจำลอง ที่ใช้พิจารณาตัวแปรสุ่ม และ prior $P(\theta)$ เพื่อคำนวณ posterior $P(\theta|\mathcal{D})$ ตามกฎของเบย์

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)}{P(\mathcal{D})}P(\theta)$$

พิจารณาให้ M แทนจำนวนของตัวแบบจำลอง Ensembles ของตัวแบบจำลองความน่าจะเป็นแต่ละตัว $\{P(y|x; \theta^{(m)})\}_{m=1}^M$ จากตัวอย่างสุ่ม Posterior $P(\theta|\mathcal{D})$ ซึ่งแต่ละตัวแบบ $P(y|x, \theta^{(m)})$ มีค่าความไม่แน่นอนของข้อมูลแตกต่างกัน จะแสดงเอนโทรปีของการแจกแจงของการทำนาย และความไม่แน่นอนในผลทำนาย เกิดจากความไม่แน่นอนของตัวแบบจำลอง ซึ่งแสดงออกมาเป็นระดับของการกระจายของตัวแบบจำลองใน Ensemble

ให้ $P(\theta|\mathcal{D})$ เป็น Predictive posterior ของ Ensembles ที่ได้มาจากค่าคาดหวังของตัวแบบจำลองใน Ensemble

$$P(y|x, \mathcal{D}) = \mathbb{E}_{p(\theta|\mathcal{D})}[P(y|x; \theta)] \approx \frac{1}{M} \sum_{m=1}^M P(y|x; \theta^{(m)}), \theta^{(m)} \sim P(\theta, \mathcal{D})$$

เอนโทรปี คือ การวัดความไม่แน่นอน หรือการสุ่มในการแจกแจงความน่าจะเป็น โดยหาปริมาณของค่าเฉลี่ยของข้อมูลที่มีอยู่ในตัวแปรสุ่ม ซึ่งพิจารณาจากความน่าจะเป็นของผลลัพธ์ที่แตกต่างกัน ถ้าค่าเอนโทรปีสูงแสดงถึงความไม่แน่นอนสูง แล้วค่าเอนโทรปีต่ำแสดงถึงความแน่นอน ดังนั้น เอนโทรปีของ Predictive posterior เป็นค่าประมาณ ความไม่แน่นอนรวม (Total Uncertainty) ซึ่งคำนวณจากสมการ

$$\mathcal{H}[P(y|x, \mathcal{D})] = \mathbb{E}_{P(y|x, \mathcal{D})}[-\ln P(y|x, \mathcal{D})]$$

ความไม่แน่นอนรวม คือผลรวมระหว่าง ความไม่แน่นอนของข้อมูล และความไม่แน่นอนของตัวแบบจำลอง อย่างไรก็ตามในการประยุกต์ใช้งานจะเหมือนกับ Active learning และ การตรวจจับนอกโดเมนจะประเมินจากค่าประมาณของ ความไม่แน่นอนของตัวแบบจำลอง โดยความไม่แน่นอนสามารถพิจารณาจาก สารสนเทศร่วม (Mutual information) ระหว่าง พารามิเตอร์ θ และ ผลทำนายของ y

Model uncertainty Total uncertainty Expected Data uncertainty

$$\begin{aligned} J[y, \theta|x, \mathcal{D}] &= \mathcal{H}[P(y|x, \mathcal{D})] - \mathbb{E}_{P(\theta|\mathcal{D})}[\mathcal{H}[P(y|x; \theta)]] \\ &\approx \mathcal{H}\left[\frac{1}{M} \sum_{m=1}^M P(y|x; \theta^{(m)})\right] - \frac{1}{M} \sum_{m=1}^M \mathcal{H}[P(y|x; \theta^{(m)})] \end{aligned}$$

จากสมการด้านบนเป็นการแสดงถึงความแตกต่างระหว่างเอนโทรปีของ Predictive Posterior (ขนาดของความไม่แน่นอนรวม) และค่าคาดหวังของเอนโทรปีของแต่ละตัวแบบจำลองใน Ensemble (ขนาดของค่าคาดหวังความไม่แน่นอนของข้อมูล) ซึ่งแตกต่างที่ขนาดของ Ensemble ที่หลากหลาย และค่าประมาณของความไม่แน่นอนของตัวแบบจำลอง

อย่างไรก็ตาม เมื่อพิจารณาตัวแบบจำลองความถดถอยความน่าจะเป็น (Probabilistic regression model) ของ Ensemble $\{P(y|x; \theta^{(m)})\}_{m=1}^M$ นอกเหนือจากความต่อเนื่องของ $y \in \mathbb{R}$ ไม่สามารถที่จะติดตามหาค่าประมาณของเอนโทรปีของ Predictive posterior (ที่แตกต่างกัน) และสารสนเทศร่วม (Mutual information) ดังนั้นในกรณีนี้ ค่าประมาณของความไม่แน่นอนสามารถหาได้จากกฎของความแปรปรวนทั้งหมด [8]

$$\mathbb{V}_{P(y|x, \mathcal{D})}[y] = \mathbb{V}_{P(\theta|\mathcal{D})}[\mathbb{E}_{P(y|x, \theta)}[y]] + \mathbb{E}_{P(\theta|\mathcal{D})}[\mathbb{V}_{P(y|x, \theta)}[y]]$$

ในวิทยานิพนธ์นี้เราใช้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอนในการศึกษาเปรียบเทียบกับนั้นเป็นอัลกอริทึมการเรียนรู้ของ Machine Learning แบบ Gradient Boosting with a Bayesian framework ซึ่งเราจะกำหนดการแจกแจงความน่าจะเป็นของ Prior $P(\theta)$ เป็นการแจกแจงแบบปกติ (Normal distribution) สำหรับตัวแปรตาม โดยมี 3 เหตุผลที่กำหนดการแจกแจงความน่าจะเป็นของ Prior เป็นการแจกแจงแบบปกติ:

1. ความสะดวกในการคำนวณ: การแจกแจงแบบปกตินั้นใช้กันอย่างแพร่หลายในสถิติแบบเบย์ (Bayesian statistics) เนื่องจากคุณสมบัติทางคณิตศาสตร์ทำให้การคำนวณ และการอนุมานเป็นไปได้ง่ายขึ้น ซึ่งจะช่วยให้กระบวนการสร้างแบบจำลอง และการคำนวณ เป็นไปอย่างมีประสิทธิภาพ นั้นง่ายขึ้น
2. ความยืดหยุ่น: ถึงแม้ว่า การแจกแจงแบบปกติอาจจะไม่สามารถแสดงถึงการแจกแจงที่แท้จริงของตัวแปรตามได้อย่างแม่นยำ แต่ก็สามารถกำหนดรูปแบบพื้นฐานของ Prior ที่ยืดหยุ่น และหลากหลายได้
3. ง่ายต่อการตีความ: การแจกแจงแบบปกติเป็นรูปแบบการแจกแจงความน่าจะเป็นที่เป็นที่รู้จัก และใช้งานกันอย่างแพร่หลายในเชิงสถิติ ซึ่งทำให้ง่ายต่อการตีความผลลัพธ์ และ

สรุปความหมาย ซึ่งค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงแบบปกติ นั้นเข้าใจ
ง่าย และแสดงข้อมูลพฤติกรรมของตัวแปรตาม

สำหรับตัวแบบจำลองความถดถอยความน่าจะเป็นของ Ensemble ซึ่งปรับพารามิเตอร์ของ
การแจกแจงแบบปกติ และโดยที่แต่ละแบบจำลองให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน สามารถ
คำนวณความแปรปรวนทั้งหมดได้ดังนี้

$$\begin{array}{ccc} \text{Total uncertainty} & \text{Model uncertainty} & \text{Expected Data} \\ \mathbb{V}_P(y|x, \mathcal{D})[y] \approx \frac{1}{M} \sum_{m=1}^M \left[\left(\sum_{m=1}^M \frac{\mu_m}{M} \right) - \mu_m \right] + \frac{1}{M} \sum_{m=1}^M \sigma_m^2 & , \{\mu_m, \sigma_m\} = f(x; \theta^{(m)}) \end{array}$$

อย่างไรก็ตาม แม้ว่าวิธีการวัดค่าเหล่านี้สามารถการคำนวณ และอ้างอิงกับค่าโมเมนต์อันดับหนึ่ง และ
อันดับสอง แต่อาจพลาดรายละเอียดของความไม่แน่นอนที่โมเมนต์อันดับสูงขึ้น นอกจากนี้ การวัด
เหล่านี้ไม่สามารถปรับเปลี่ยนขนาดได้อิสระ ซึ่งอาจทำให้เกิดปัญหาเมื่อมีค่าการคาดการณ์ที่มีขนาด
ที่แตกต่างกันระหว่างข้อมูลในโดเมนและข้อมูลนอกโดเมน [8] [9]

2.6 ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

2.6.1 NGBoost (Natural Gradient Boosting)

เป็นอัลกอริทึมประเภท การเรียนรู้แบบมีผู้สอน (Supervised Learning) สำหรับการทำนายความน่าจะเป็นทั่วไปโดยใช้วิธี Gradient Boosting เพื่อประมาณพารามิเตอร์ของการ แจกแจงความน่าจะเป็นแบบมีเงื่อนไข $P(y|x)$ เป็นฟังก์ชันของ x โดยตัวแบบจำลองถดถอยทั่วไปจะส่งค่าประมาณแบบจุดอย่างมีเงื่อนไข แต่ตัวแบบจำลองการถดถอยความน่าจะเป็นจะส่งการแจกแจงความน่าจะเป็นของผลลัพธ์ ซึ่งช่วยให้สามารถประมาณค่าความไม่แน่นอนเชิงคาดการณ์ได้แบบมีเงื่อนไขได้ ซึ่งมีความสำคัญในการใช้งานด้านการแพทย์ และด้านพยากรณ์อากาศ ซึ่งการถดถอยความน่าจะเป็นนี้ขึ้นอยู่กับพารามิเตอร์ของการแจกแจงแบบมีเงื่อนไขเป็นเป้าหมายสำหรับอัลกอริทึม Boosting หลายพารามิเตอร์ θ

สำหรับการเรียนรู้ในการพยากรณ์แบบจุด ในการพยากรณ์การถดถอยความน่าจะเป็นโดยใช้ฟังก์ชันการสูญเสีย (Loss function) ซึ่งต้องมีเกณฑ์คะแนนที่ต้องเปรียบเทียบการประมาณการแจกแจงความน่าจะเป็นของข้อมูล โดยใช้เกณฑ์คะแนนที่เหมาะสม (Proper scoring rule) \mathcal{S} โดยใช้ข้อมูลการแจกแจงความน่าจะเป็นที่ถูกพยากรณ์ P และค่าสังเกต y แล้วกำหนด $\mathcal{S}(P; y)$ กับการพยากรณ์เพื่อให้การแจกแจงจริงของผลลัพธ์ได้คะแนนที่ดีที่สุดโดยคาดหวัง

ในสัญกรณ์ทางคณิตศาสตร์, Proper scoring rule \mathcal{S} เป็นเกณฑ์คะแนนที่เหมาะสม ก็ต่อเมื่อ เป็นไปตามเงื่อนไข

$$\mathbb{E}_{y \sim Q}[\mathcal{S}(Q, y)] \leq \mathbb{E}_{y \sim Q}[\mathcal{S}(P, y)] \quad \forall P, Q$$

กำหนดให้ Q เป็นการแจกแจงจริงของผลลัพธ์ของ y และ P เป็นการแจกแจงอื่น ๆ โดยที่เราใช้การแจกแจงพารามेटริก (Parametric distribution) ในการทำงาน จึงเราสามารถระบุแต่ละการแจกแจงพารามิเตอร์ และเกณฑ์คะแนนเป็น $\mathcal{S}(\theta, y)$ [9]

2.6.2 XGBoost-Distribution (eXtreme Gradient Boosting Distribution)

เป็นอัลกอริทึมประเภท การเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งมีตัวแบบจำลอง XGBoost เป็นพื้นฐานอัลกอริทึมการ Gradient Boosting กับต้นไม้ตัดสินใจ โดยการหาค่าที่เหมาะสมที่สุด เพื่อเพิ่มประสิทธิภาพในการทำงาน ใช้เทคนิค Weighted quantile sketch เพื่อค้นหาจุดแบ่งช่วงที่เป็นไปได้ในข้อมูลอย่างมีประสิทธิภาพ นอกจากนี้ยังรองรับการแบ่งการสร้างต้นไม้แบบพร้อมกันเพื่อเพิ่มความเร็วในการฝึก ซึ่งกระจายการคำนวณในหลายหน่วยประมวลผล (CPU

cores) เพื่อความสำเร็จและความแม่นยำสูงในงานจำแนกหรือทำนาย ในขณะที่สามารถจัดการกับความซับซ้อนของโมเดลและปัญหาการเหลื่อมหรือติดได้อย่างดี โดยให้ความสำคัญของตัวแปร ซึ่งแสดงถึงความสำคัญที่สูงต่อแต่ละตัวแปรในการพยากรณ์ การคำนวณความสำคัญของตัวแปรใช้ความถี่ของการใช้ตัวแปรในการแบ่งช่วงข้อมูลระหว่างต้นไม้ทั้งหมดในต้นไม้แบบจำลอง และการปรับปรุงเฉลี่ยในฟังก์ชันสูญเสีย โดยฟังก์ชันสูญเสียของ XGBoost-Distribution ที่ใช้ในการพยากรณ์การถดถอยความน่าจะเป็น คือ proper scoring rule \mathcal{S} ซึ่งเป็นฟังก์ชันการสูญเสียเดียวกับ NGBoost [10]

2.6.3 Catboost (Gradient Boosting with categorical features support)

Catboost มาจากคำว่า “Category” และ “Boosting” เป็นอัลกอริทึมประเภท การเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้วิธี Gradient Boosting ที่ขึ้นกับวิธีต้นไม้การตัดสินใจ การเรียนรู้ของเครื่องจาก Yandex เปิดให้ใช้งานแบบโอเพนซอร์ส Catboost สามารถใช้งานร่วมกับเฟรมเวิร์กการเรียนรู้เชิงลึก ซึ่งจุดเด่นของ CatBoost คือการใช้อัลกอริทึมที่พิเศษในการจัดการคุณลักษณะทางสถิติแบบหมวดหมู่ หรือเรียกว่า Ordered Target Statistics (OTS) เพื่อทำการเตรียมข้อมูลทางสถิติสำหรับคุณลักษณะทางสถิติแบบหมวดหมู่ โดยการคำนวณสถิติสำหรับแต่ละหมวดหมู่ การใช้วิธีนี้ช่วยให้จัดการคุณลักษณะทางสถิติแบบหมวดหมู่ได้อย่างมีประสิทธิภาพโดยไม่ต้องทำการเตรียมข้อมูลหรือการเข้ารหัสแบบหมวดหมู่เป็นหมายเลขให้ยาว รวมถึงโครงสร้างต้นไม้ที่สมมาตร (Symmetric Trees) ซึ่งมีโครงสร้างเดียวกันในส่วนของกิ่งทางซ้ายและกิ่งทางขวา สิ่งนี้ช่วยให้สามารถประมวลผลได้เร็วขึ้น และฝึกได้อย่างรวดเร็ว นอกจากนี้มีความสามารถในการปรับขนาดได้อย่างยอดเยี่ยม

ซึ่งถูกออกแบบให้มีความเร็วและสามารถปรับขนาดได้ มันใช้เทคนิคการแบ่งงานแบบพร้อมกันเพื่อประมวลผลชุดข้อมูลที่ใหญ่อย่างมีประสิทธิภาพ รวมถึงการรองรับฟังก์ชันสูญเสียพิเศษสำหรับงานที่แตกต่างกัน โดยฟังก์ชันสูญเสียพิเศษสำหรับการพยากรณ์การถดถอยความน่าจะเป็นดังนี้

$$-\sum_{i=1}^N \log N(t_i | a_{i,0}, e^{2a_{i,1}}) = \frac{1}{2} \log(2\pi) + \sum_{i=1}^N \left(a_{i,1} + \frac{1}{2} e^{-2a_{i,1}} (t_i - a_{i,0})^2 \right)$$

เมื่อ t คือตัวแปรตาม โดย a_0 เป็นการประมาณเชิงสถิติ 2 มิติของการพยากรณ์ตัวแปรตาม และ a_1 เป็นการพยากรณ์เชิงสถิติ $\log \sigma$ และ $N(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)}$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นที่เป็นไปตามการแจกแจงปกติ [8] [11]

2.7 การเทียบมาตรฐาน และการปรับเทียบมาตรฐานสำหรับแบบจำลองการถดถอย

โดยสมมติให้ $\mathcal{Y} = \{0,1\}$ เราจะกำหนดชุดข้อมูลให้ $x_t, y_t \in \mathcal{X} \times \mathcal{Y}$ สำหรับ $t = 1, 2, \dots, T$ โดยตัวแปรสุ่ม \mathcal{X} และตัวแปรสุ่ม \mathcal{Y} ที่มีการแจกแจงเดียวกันและเป็นอิสระกันต่อกัน (i.i.d.) ที่ถูกสุ่มมาจากการแจกแจงความน่าจะเป็น \mathbb{P} เราจะให้ตัวพยากรณ์ (Forecaster) $H: \mathcal{X} \rightarrow (\mathcal{Y} \rightarrow [0,1])$ เพื่อสร้างการแจกแจงความน่าจะเป็น $F_t(y)$ ซึ่งเป็นฟังก์ชันการแจกแจงความน่าจะเป็นสะสม (CDF) เมื่อ Y เป็นตัวแปรต่อเนื่อง ที่เน้นการให้เป้าหมาย y_t โดยมีเงื่อนไขว่า x_t เป็นตัวกำหนด

วิธีการเทียบมาตรฐาน (Calibration) และการปรับเทียบมาตรฐาน (Recalibration) เพื่อแก้ไขปัญหของแบบจำลองการถดถอยที่ให้ค่าประมาณ $\mathcal{Y} = \mathbb{R}$ และนำอัลกอริทึมที่ได้มาใช้กับตัวแบบจำลองความไม่แน่นอน ซึ่ง Forecaster H ในแต่ละ t สร้าง CDF แล้ว F_t ซึ่งพยากรณ์ y_t เมื่อ $F_t^{-1}(p)$ กำหนดให้เป็นค่าที่ต่ำที่สุดของ y ที่ $p \leq F_t(y)$ แล้ว ให้ฟังก์ชัน $F_t^{-1}: [0,1] \rightarrow \mathcal{Y}$ แสดงฟังก์ชันควอไทล์ (Quantile)

2.7.1 การเทียบมาตรฐานสำหรับแบบจำลองการถดถอย (Calibration Regression)

คือ การวัดระดับความเชื่อมั่นของการทำนาย โดยที่ผลลัพธ์จริง y_t อยู่ช่วงความเชื่อมั่นของผลทำนายที่กำหนดไว้หรือไม่ โดยมีเป้าหมายเพื่อให้มีการทำนายค่าที่ตรงกับช่วงความเชื่อมั่นนั้นๆ ในระดับความเชื่อมั่นที่กำหนดไว้ โดย ตัวพยากรณ์ (Forecaster) H จะถูกพิจารณาว่า Calibrated ถ้าฟังก์ชันการแจกแจงความน่าจะเป็นสะสมเชิงประจักษ์ และเชิงทำนาย (CDFs) มีความสอดคล้องกับค่าจริง เมื่อขนาดของชุดข้อมูลเข้าสู่ความไม่จำกัด นิยามได้ดังนี้ [12]

$$\frac{\sum_{t=1}^T \mathbb{I}\{y_t \leq F_t^{-1}(p)\}}{T} \rightarrow p \text{ for all } p \in [0,1]$$

เมื่อ พิจารณา x_t และ y_t เป็นตัวอย่างที่สุ่มมาแบบอิสระกันและ i.i.d. ระหว่างแปรสุ่ม X และ Y ตามลำดับ โดยการแจกแจงความน่าจะเป็น \mathbb{P} ดังนั้นมีเงื่อนไขเพียงพอสำหรับ Calibration คือ

$$\mathbb{P}(Y \leq F_X^{-1}(p)) = p \text{ for all } p \in [0,1]$$

โดยที่ $F_X = H(X)$ หมายถึง การพยากรณ์ที่ X

จากนั้น ด้วยการใช้เทคนิคแบบเบย์ เช่น เครือข่ายประสาทแบบเบย์ (Bayesian neural networks) หรือกระบวนการเกาส์เซียน (Gaussian processes) สร้างการพยากรณ์ด้วยเทคนิคเชิงความน่าจะเป็น F_t โดยที่ เราจะเลือก m ระดับความเชื่อมั่น $0 \leq p_1 < p_2 < \dots < p_m \leq 1$ และ คำนวณความถี่เชิงประจักษ์สำหรับแต่ละค่า p_j เพื่อแสดงผลการสอดคล้องกันของการพยากรณ์

$$\hat{p}_j = \frac{|\{y_t | F_t(y_t) < p_j, t = 1, 2, \dots, T\}|}{T}$$

สุดท้ายแล้ว สำหรับเกณฑ์คะแนนที่วัดประสิทธิภาพของการทำนายตัวแบบ เราจะใช้ ความผิดพลาดของการเทียบมาตรฐาน (Miscalibration Area)

$$cal(F_1, y_1, \dots, F_T, y_T) = \sum_{j=1}^m (p_j - \hat{p}_j)^2$$

2.7.2 การปรับเทียบมาตรฐานสำหรับแบบจำลองการถดถอย (Recalibration Regression)

กระบวนการนี้เกี่ยวกับการประมาณค่าความน่าจะเป็นที่แท้จริง $\mathbb{P}(Y \leq F_X^{-1}(p))$ ของตัวแปรสุ่ม Y ที่อยู่ภายในบริเวณ $(-\infty, F_X^{-1}(p)]$ สำหรับระดับความเชื่อมั่นใดๆ โดยสามารถกำหนดให้ทุกๆ การพยากรณ์ F_t เป็น $G \circ F_t$ โดยที่ $G(p) := \mathbb{P}(Y \leq F_X^{-1}(p))$ จะได้ Forecaster ที่มี Calibrated อย่างสมบูรณ์แบบ (Well-calibration) ซึ่ง วัตถุประสงค์ของ Recalibration คือการประมาณการแจกแจงความน่าจะเป็นทั้งหมดโดยใช้ชุดข้อมูล $\{(x_t, y_t)\}_{t=1}^T$ ซึ่งสามารถทำการปรับขั้นตอนของแบบจำลองการถดถอยกลับให้เข้ากับชุด Recalibration $\{F_t(y_t), \hat{P}(F_t(y_t))\}_{t=1}^T$:

$$\hat{P}(p) = \frac{|\{y_t | F_t(y_t) < p, t = 1, 2, \dots, T\}|}{T}$$

โดยที่ $\hat{P}(p)$ เป็นสัดส่วนของจุดข้อมูลที่ y_t ต่ำกว่า Quantile ลำดับที่ p ของ F_t ซึ่งในวิทยานิพนธ์นี้ ใช้การถดถอยแบบไอโซโทนิก เป็นซึ่งเป็นวิธี Non-Parametric และสามารถประมาณค่าการแจกแจงที่แท้จริงได้อย่างแม่นยำ เมื่อมีข้อมูลที่เป็น i.i.d. เพียงพอ ซึ่งวิธีนี้อธิบายถึงการเปลี่ยนแปลงของฟังก์ชันจริง $\mathbb{P}(Y \leq F_X^{-1}(p))$ ซึ่งเพิ่มขึ้นแบบโมโนโทนิก (คือฟังก์ชันหนึ่งทีค่าเพิ่มขึ้นหรือคงเดิมเมื่อค่าตัวแปรเพิ่มขึ้น ในขณะที่ฟังก์ชันลดระดับโมโนโทนิกคือฟังก์ชันหนึ่งที่ค่าลดลงหรือคงเดิมเมื่อค่าตัวแปรเพิ่มขึ้น) กล่าวคือ จะได้ Forecaster ที่มี Calibrated เป็น Well-calibration [12]

2.8 วิธีการแก้ปัญหาการลดเวลาการทดสอบ

2.8.1 แนวคิดสำหรับแก้ปัญหาการทดสอบกระแสเขียน

การทดสอบกระแสเขียนเกี่ยวข้องกับการส่งตัวฮาร์ดไดรฟ์ไปยังช่วงค่ากระแสเขียน และการวัดพารามิเตอร์บางอย่างสำหรับแต่ละค่ากระแสเขียน เช่น ความสามารถในการเลื่อนตำแหน่ง (Tracks-Per-Inch capability) ความสามารถในการบันทึกข้อมูล (Bits-Per-Inch capability) ลักษณะการรบกวนระหว่างแทร็กที่อยู่ใกล้เคียง (Adjacent-Track Interference characteristics) และอื่น ๆ โดยอิงจากพารามิเตอร์ที่วัดได้ จากพารามิเตอร์ที่วัดได้ เลือกกระแสเขียนที่เหมาะสมที่สุดที่จะใช้ในการทำงานของฮาร์ดไดรฟ์ การทดสอบกระแสเขียนนี้มักทำหลังจากเก็บรวบรวมข้อมูลจำนวนมากเกี่ยวกับฮาร์ดไดรฟ์จากการทดสอบก่อนหน้านี้ ซึ่งอาจรวมถึงค่าความต้านทานขององค์ประกอบแม่เหล็ก-ต้านทานแม่เหล็กในหัวอ่าน/เขียน (Read/Write head's magneto-resistive element) และค่ากระแสความร้อนของอ่านข้อมูล (Reader heat current)

เพื่อกำหนดชุดค่ากระแสเขียนที่จะทดสอบให้มีจำนวนเล็กน้อย ในงานนี้ใช้โครงข่ายประสาทเทียม (Artificial Neural Network) โดยโครงข่ายประสาทเทียมถูกฝึกสอน โดยใช้ตัวอย่างที่รวมถึงข้อมูลการทดสอบก่อนหน้าที่เกี่ยวข้องกับแต่ละฮาร์ดไดรฟ์ โครงข่ายประสาทเทียมคำนวณรูปแบบที่กำหนดช่วงของค่ากระแสเขียนที่ควรทดสอบ เพื่อตั้งค่ากระแสเขียนของแต่ละฮาร์ดไดรฟ์ จากนั้นฮาร์ดไดรฟ์แบบสองของชุดที่สองจะถูกส่งผ่านการทดสอบกระแสเขียนแบบจำกัด เพื่อตั้งค่ากระแสเขียนที่ใช้งานในแต่ละฮาร์ดไดรฟ์

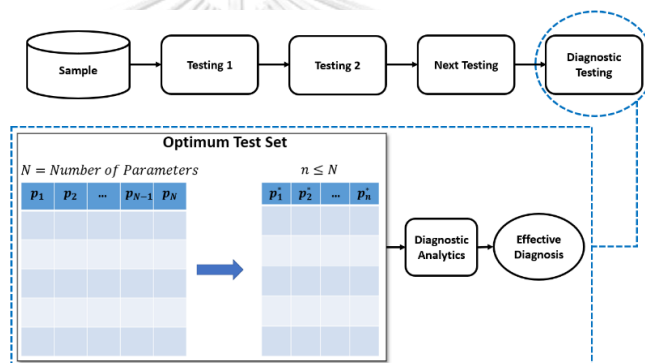
ในบางกรณีที่ระบุไว้ในงานนี้ ชุดแรกของฮาร์ดไดรฟ์ที่ใช้ในการฝึกสอนจะเป็นส่วนน้อยของจำนวนรวมของฮาร์ดไดรฟ์ที่ผลิตขึ้น และจะต้องเป็นตัวแทนของฮาร์ดไดรฟ์ที่จะทดสอบในภายหลัง ข้อมูลการฝึกสอนสามารถใช้ในการพยากรณ์ปรับช่วงสำหรับแต่ละรุ่นฮาร์ดไดรฟ์ หรือชุดเดียวของฮาร์ดดิสก์ไดรฟ์ โครงข่ายประสาทเทียมสามารถถูกฝึกสอนใหม่ เมื่อมีการเปลี่ยนแปลงกระบวนการผลิตบางส่วนซึ่งอาจส่งผลต่อความเป็นตัวแบบจำลองที่ใช้ของข้อมูลการฝึกสอนเดิมสำหรับฮาร์ดไดรฟ์ที่ผลิตภายใต้กระบวนการที่แตกต่างกัน โดยโครงข่ายประสาทเทียม และส่วนประกอบของมันสามารถนำมาใช้งานในอุปกรณ์คอมพิวเตอร์หนึ่ง หรือมากกว่านั้นที่มีตัวประมวลผลและหน่วยความจำ และวิธีการสามารถดำเนินการโดยใช้คำสั่ง/โค้ดที่เขียนอยู่ในหน่วยความจำซึ่งสามารถอ่านได้โดยเครื่องคอมพิวเตอร์ [2]

ถึงแม้อย่างไรก็ตามแนวคิดในการแก้ปัญหาการทดสอบกระแสเขียนนี้ ยังไม่ผลลัพธ์ หรืองานวิจัยที่ชี้ให้เห็นว่าแนวคิดนี้สามารถแก้ปัญหาในการทดสอบกระแสเขียนได้ นอกจากนี้ในการติดตั้งตัวแบบจำลองประเภทโครงข่ายประสาทเทียมในเครื่องทดสอบไม่ได้มีฮาร์ดแวร์สูง ซึ่งไม่สามารถรองรับการประมวลผลของตัวแบบจำลองประเภทโครงข่ายประสาทเทียม

2.8.2 เงื่อนไขการทดสอบแบบปรับตัว (Adaptive Test Condition)

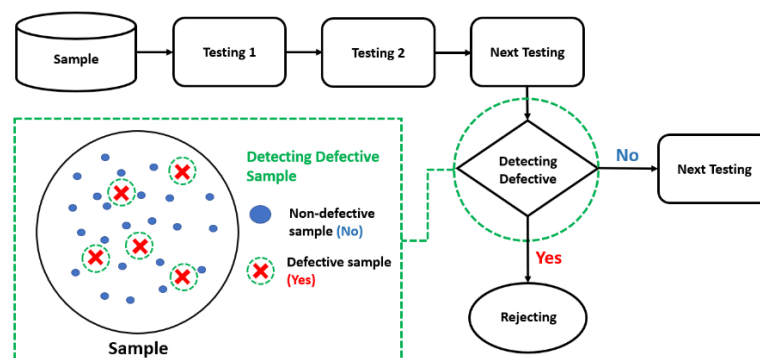
เป็นเทคนิคที่ปรับปรุงวิธีการทดสอบดั้งเดิม โดยที่ปรับเปลี่ยนการทดสอบไปตามรูปแบบ และความสัมพันธ์ของพารามิเตอร์ เพื่อเพิ่มประสิทธิภาพในการทดสอบ [3] เช่น สามารถระบุหน่วยตัวอย่างที่ผิดปกติได้ถูกต้องมากขึ้น ลดเวลาในการทดสอบลง เป็นต้น โดย Adaptive Test Condition ที่ใช้ปรับปรุงการลดเวลาในการทดสอบมีอยู่ด้วยหัน 3 วิธี ดังนี้

1. การเลือกชุดทดสอบที่เหมาะสมที่สุด (Optimal test set) จากรูปที่ 8 การค้นหาชุดทดสอบที่ดีที่สุดเพื่อปรับปรุงประสิทธิภาพในการตรวจวินิจฉัยและลดเวลาทดสอบ โดยการเลือกพารามิเตอร์ (p^*) จากทั้งหมดของพารามิเตอร์ (p) ที่สำคัญสำหรับการวินิจฉัย ซึ่งกำหนดรูปแบบจากการทดสอบ จากพารามิเตอร์ที่มีสัมพันธ์กัน และการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการวินิจฉัย [13] [14] [15] [16]



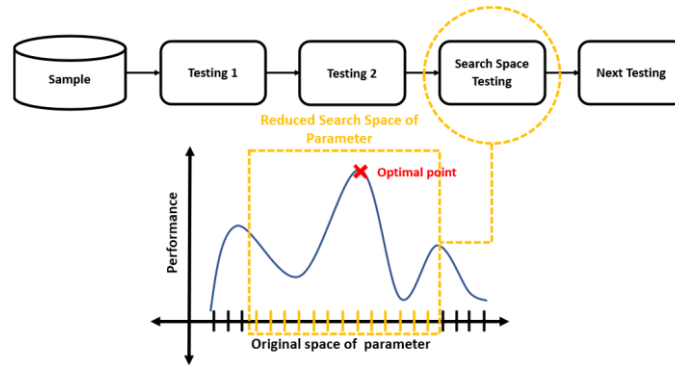
รูปที่ 8 การเลือกชุดทดสอบที่เหมาะสมที่สุด (Optimal test set)

2. การตรวจจับตัวอย่างที่มีข้อบกพร่อง (Detecting defective samples) จากรูปที่ 9 การตรวจจับข้อบกพร่องเพื่อลดจำนวนตัวอย่างที่ต้องทดสอบก่อนเข้าสู่กระบวนการทดสอบถัดไป โดยวิธีการตัดสินใจจากแบบจำลองที่ได้จากข้อมูลการทดสอบพารามิเตอร์ และพารามิเตอร์อื่นๆ ที่มีความสัมพันธ์กันกับการวินิจฉัย [17] [18] [19] [20]



รูปที่ 9 การตรวจจับตัวอย่างที่มีข้อบกพร่อง (Detecting defective samples)

3. การลดช่วงค้นหาในการทดสอบของพารามิเตอร์ (Reducing search space of parameter) จากรูปที่ 10 นำวิธีการทางสถิติใช้กับพารามิเตอร์ที่เกี่ยวข้องกับการทดสอบพารามิเตอร์นั้น แล้วนำผลวิเคราะห์ที่ได้มาลดพื้นที่การทดสอบค้นหา [21]



รูปที่ 10 การลดช่วงค้นหาในการทดสอบของพารามิเตอร์ (Reducing search space of parameter)

2.9 เกณฑ์คะแนนที่ใช้ในการศึกษา และเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

2.9.1 MAE (Mean Absolute Error)

คือ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งเป็นวิธีการวัดค่าความคลาดเคลื่อนที่นิยม โดยจะช่วยบอกถึงขนาดของความคลาดเคลื่อนรวมเฉลี่ยได้ ยิ่งค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยมีค่าน้อยแสดงว่า ตัวแบบจำลองที่ได้จะมีความแม่นยำสูง

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

โดยที่

n คือ จำนวนตัวอย่าง

y_i คือ ค่าจริง ลำดับที่ i

\hat{y}_i คือ ค่าจริงที่ได้จากการพยากรณ์ของตัวแบบ ลำดับที่ i

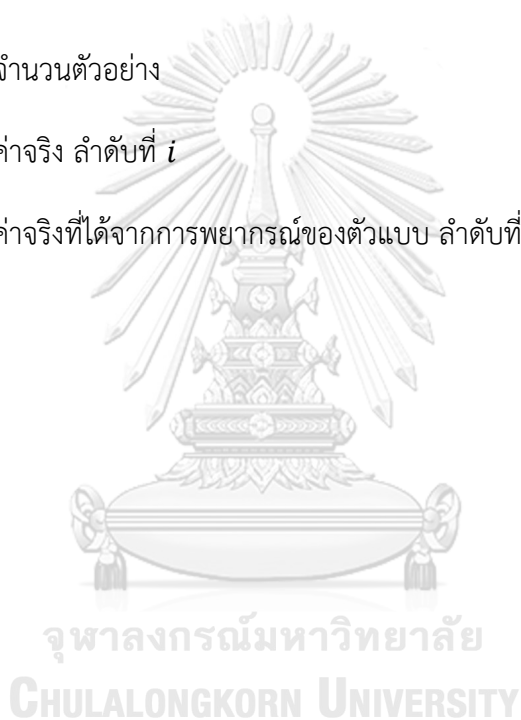
2.9.2 RMSE (Root Mean Squared Error)

คือ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งเป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐานที่นิยม ยิ่งค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าน้อย แสดงว่า ตัวแบบจำลองที่ได้จะมีความแม่นยำสูง

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

โดยที่

- n คือ จำนวนตัวอย่าง
- y_i คือ ค่าจริง ลำดับที่ i
- \hat{y}_i คือ ค่าจริงที่ได้จากการพยากรณ์ของตัวแบบ ลำดับที่ i



2.9.3 AvgTTR

คือ การประมาณเวลาที่ลดลงจากการทดสอบกระแสไฟฟ้าเขียน โดยเกณฑ์คะแนนนี้ใช้ในการเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่แน่นอน โดยต้องครอบคลุมสัดส่วนชุดตัวอย่างอย่างน้อย 0.9

$$AvgTTR = n_{head} \times n_{zone} \times \left(\frac{\sum_{i=1}^n \Delta range_i}{n} \times RunTime \right)$$

โดย

$$cl_{min} = arg \min_{\alpha \in [0,1]} \{P(y_i \in [\hat{L}_{cl_i}, \hat{U}_{cl_i}]) > 0.9\}$$

$$(\hat{U}_{cl_{min_i}}, \hat{L}_{cl_{min_i}}) = \{\hat{y}_i \in \hat{Y}, \sigma_i \in \sigma | (\hat{y}_i - Z_{cl_{min}} \sigma_i, \hat{y}_i + Z_{cl_{min}} \sigma_i)\}$$

$$U_{update_i} = \begin{cases} \min_{i \in n} (U_i, \hat{U}_{cl_{min_i}}), & y_i \in [\hat{L}_{cl_{min_i}}, \hat{U}_{cl_{min_i}}] \\ U_i, & \text{other} \end{cases}$$

$$L_{update_i} = \begin{cases} \max_{i \in n} (L_i, \hat{L}_{cl_{min_i}}), & y_i \in [\hat{L}_{cl_{min_i}}, \hat{U}_{cl_{min_i}}] \\ L_i, & \text{other} \end{cases}$$

$$\Delta range_i = (U_i - L_i) - (U_{update_i} - L_{update_i})$$

เมื่อ	n	คือ จำนวนตัวอย่าง
	n_{head}	คือ จำนวนหัวเขียน
	n_{zone}	คือ จำนวนโซนบนแผ่นดิสก์ของฮาร์ดไดรฟ์
	$RunTime$	คือ เวลาเฉลี่ยในการวัดค่าระยะ 1 หน่วย
	σ_i	คือ ค่าความไม่แน่นอนของข้อมูล σ จากตัวแบบจำลองการถดถอยโดยความไม่แน่นอนที่ผ่านการปรับการเทียบมาตรฐาน ลำดับที่ i
	cl	คือ ระดับความเชื่อมั่น
	U_i	คือ ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้าเขียนวัดได้จากการทดสอบฮาร์ดไดรฟ์ ลำดับที่ i
	L_i	คือ ขอบเขตล่างของช่วงการวัดค่ากระแสไฟฟ้าเขียนวัดได้จากการทดสอบฮาร์ดไดรฟ์ ลำดับที่ i

\hat{U}_{cl_i} คือ ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้า
เขียนที่ได้จากตัวแบบจำลองการถดถอยโดย
ความไม่แน่นอน ลำดับที่ i
ณ ระดับความเชื่อมั่นที่ cl

\hat{L}_{cl_i} คือ ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้า
เขียนที่ได้จากตัวแบบจำลองการถดถอยโดย
ความไม่แน่นอน ลำดับที่ i ณ ระดับความเชื่อมั่น
ที่ cl

cl_{min} คือ ระดับความเชื่อมั่นที่น้อยที่สุดในช่วง
[0,1) ที่ทำให้ความน่าจะเป็นที่ $y_i \in$
[$\hat{L}_{cl_i}, \hat{U}_{cl_i}$] ณ ระดับความเชื่อมั่นที่ cl มี
มากกว่า 0.9

$\hat{U}_{cl_{min}i}$ คือ ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้า
เขียนที่ได้จากตัวแบบจำลองการถดถอยโดย
ความไม่แน่นอน ลำดับที่ i ณ ระดับความเชื่อมั่น
ที่ cl_{min}

$\hat{L}_{cl_{min}i}$ คือ ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้า
เขียนที่ได้จากตัวแบบจำลองการถดถอยโดย
ความไม่แน่นอน ลำดับที่ i ณ ระดับความเชื่อมั่น
ที่ cl_{min}

U_{update_i} คือ ถ้า $y_i \in [\hat{L}_{cl_{min_i}}, \hat{U}_{cl_{min_i}}]$ ณ ระดับความเชื่อมั่นที่ cl_{min} แล้ว ขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้าเขียนที่น้อยที่สุด ระหว่างขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้าเขียนวัดได้ โดยการจ่ายกระแสไฟฟ้า ณ ค่าคงที่หนึ่ง และขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้าเขียนที่ได้จากตัวแบบจำลองการถดถอยโดยความไม่แน่นอน ณ ระดับความเชื่อมั่นที่ cl_{min}

ลำดับที่ i ถ้าไม่เท่ากับขอบเขตบนของช่วงการวัดค่ากระแสไฟฟ้าสำหรับเขียนวัดได้จากการทดสอบฮาร์ดแวร์ ลำดับที่ i

L_{update_i} คือ ถ้า $y_i \in [\hat{L}_{cl_{min_i}}, \hat{U}_{cl_{min_i}}]$ ณ ระดับความเชื่อมั่นที่ cl_{min} แล้ว ขอบเขตล่างของช่วงการวัดค่ากระแสไฟฟ้าเขียนที่มากที่สุด ระหว่างขอบเขตล่างของช่วงการวัดค่ากระแสไฟฟ้าเขียนวัดได้ โดยการจ่ายกระแสไฟฟ้า ณ ค่าคงที่หนึ่ง และขอบเขตล่างของช่วงการวัดค่ากระแสไฟฟ้าเขียนที่ได้จากตัวแบบจำลองการถดถอยโดยความไม่แน่นอน ณ ระดับความเชื่อมั่นที่ cl_{min}

ลำดับที่ i ถ้าไม่เท่ากับขอบเขตล่างของช่วงการวัดค่ากระแสไฟฟ้าสำหรับเขียนวัดได้จากการทดสอบฮาร์ดแวร์ ลำดับที่ i

$\Delta range_i$ คือ ผลต่างระหว่างระยะความกว้างของช่วงการวัดค่า $[L_i, U_i]$ และระยะความกว้างของช่วงการวัดค่า $[L_{update_i}, U_{update_i}]$ ลำดับที่ i

$AvgTTR$ คือ เวลาในการวัดค่าค้นหากระแสไฟฟ้าเขียนที่เหมาะสมที่สุด ที่ลดลงได้โดยเฉลี่ยต่อฮาร์ดแวร์ (นาทีก)

บทที่ 3 วิธีการดำเนินงานวิจัย

3.1 ขอบเขตของงานวิจัย

ศึกษาการเปรียบเทียบวิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอน เพื่อลดเวลาในการวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์ ระหว่างวิธี CatBoost, XGBoost-Distribution และ NGBoost โดยใช้ข้อมูลประสิทธิภาพของส่วนประกอบฮาร์ดไดรฟ์ ข้อมูลประสิทธิภาพหลังประกอบเป็นฮาร์ดไดรฟ์ และข้อมูลช่วงกระแสไฟฟ้าเขียนที่วัดได้ที่ได้จากการทดสอบฮาร์ดไดรฟ์ แล้วนำค่าความไม่แน่นอนของข้อมูลที่ได้จากการเรียนรู้ของตัวแบบมา Recalibrated เพื่อให้เหมาะสม แล้วนำไปคำนวณช่วงความเชื่อมั่นของผลทำนายแต่ละระดับช่วงเชื่อมั่น แล้วจึงนำมาเปรียบเทียบความสามารถในการลดเวลาในการทดสอบกระแสไฟฟ้าเขียน โดยใช้เกณฑ์ *AvgTTR*

3.2 วิธีการดำเนินการวิจัย

3.2.1 นำเข้าข้อมูล และทำความสะอาดข้อมูลสำหรับเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

ขอขอบคุณ บริษัทซีเกทเทคโนโลยี (ประเทศไทย) จำกัด นครราชสีมา ประเทศไทย ที่ให้การสนับสนุนสถานที่ฝึกงาน และข้อมูลในการทำวิทยานิพนธ์ ซึ่งเป็นข้อมูลตัวอย่างในกระบวนการผลิตฮาร์ดไดรฟ์ นำมาวิเคราะห์ และเตรียมสำหรับการเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน เลือกเฉพาะข้อมูลตัวอย่างฮาร์ดไดรฟ์ที่ผลิตจากวัสดุประเภท A โดยแต่ละหน่วยตัวอย่างมีหน่วยเป็น แต่ละฮาร์ดไดรฟ์ต่อหัวเขียนต่อโซน ซึ่งมีจำนวนตัวอย่าง 2,382,696 โซน และจำนวนรายการทดสอบฮาร์ดไดรฟ์ 69 รายการ หลังจากนั้นเราจะตรวจสอบชุดข้อมูลชุดนี้มีหน่วยตัวอย่างที่มีข้อมูลสูญหายหรือไม่ ซึ่งเราพบว่าในชุดข้อมูลมีหน่วยตัวอย่างที่มีข้อมูลสูญหายจำนวน 34,236 โซน เราจะดำเนินการตัดทิ้งหน่วยตัวอย่างที่มีข้อมูลสูญหายทั้งหมด เนื่องจากหน่วยตัวอย่างที่มีข้อมูลสูญหายนั้นเกิดจากการเก็บข้อมูลที่ผิดพลาดของอุปกรณ์ และกระบวนการทดสอบ หลังจากตัดทิ้งหน่วยตัวอย่างที่มีข้อมูลสูญหายทั้งหมดแล้ว ชุดข้อมูลนี้จะเหลือจำนวนตัวอย่าง 2,348,460 โซน

ตารางที่ 1 แสดงชื่อตัวแปร ความหมายของตัวแปร และประเภทตัวอย่างของตัวแปร แต่ละตัวอย่าง

ตัวแปร	ความหมายของตัวแปร	ประเภทตัวอย่างของตัวแปร
week	สัปดาห์ที่ทดสอบหลังประกอบฮาร์ดไดรฟ์	Object
B1	ข้อมูลการทดสอบของแต่ละส่วนประกอบฮาร์ดไดรฟ์ก่อนประกอบเป็นฮาร์ดไดรฟ์	Float64
B2		
B3		
B4		
B5		
B6		
B7		
A1	ข้อมูลการทดสอบของฮาร์ดไดรฟ์หลังประกอบ	Float64
A2		
A3		
A4		
A5		
A6		
MIN	ค่ากระแสไฟฟ้าเขี่ยนที่น้อยที่สุดในช่วงการวัดค่ากระแสไฟฟ้าเขี่ยน สำหรับการทดสอบกระแสไฟฟ้าเขี่ยน	Int64
CURR	ค่ากลางของกระแสไฟฟ้าเขี่ยนของช่วงการวัดค่ากระแสไฟฟ้าเขี่ยน สำหรับการทดสอบกระแสไฟฟ้าเขี่ยน	
MAX	ค่ากระแสไฟฟ้าเขี่ยนที่มากที่สุดในช่วงการวัดค่ากระแสไฟฟ้าเขี่ยน สำหรับการทดสอบกระแสไฟฟ้าเขี่ยน	
Zone	โซนของแผ่นดิสก์บนฮาร์ดไดรฟ์	Category
Target	ค่ากระแสไฟฟ้าเขี่ยนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์	Int64

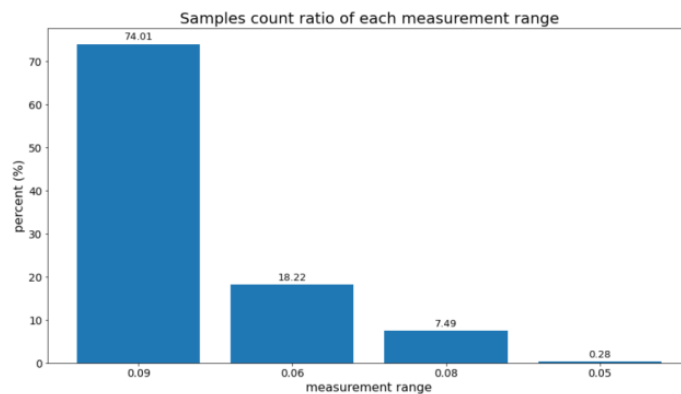
ตารางที่ 1 จะเลือกคอลัมน์สัปดาห์ที่ทดสอบหลังประกอบฮาร์ดไดรฟ์ คือ week ตัวแปรอิสระ และตัวแปรตาม ตามที่กำหนดซึ่งตัวแปรอิสระคือ รายการทดสอบฮาร์ดไดรฟ์ประกอบไปด้วยข้อมูลการทดสอบของแต่ละส่วนประกอบฮาร์ดไดรฟ์ก่อนประกอบเป็นฮาร์ดไดรฟ์ ได้แก่ B1, B2, B3,

B4, B5, B6, B7 ข้อมูลการทดสอบของฮาร์ดไดรฟ์หลังประกอบ ได้แก่ A1, A2, A3, A4, A5, A6 ข้อมูลช่วงการวัดค่ากระแสไฟฟ้าเขียนที่วัดได้ที่ได้จากการทดสอบฮาร์ดไดรฟ์ เพื่อค้นหาช่วงกระแสไฟฟ้าสำหรับการเขียนที่ส่งผลให้ฮาร์ดไดรฟ์สามารถทำงานได้ ได้แก่ MIN, CURR, MAX ข้อมูลโซนของแผ่นดิสก์บนฮาร์ดไดรฟ์ คือ Zone และตัวแปรตามคือ กระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์ คือ Target ซึ่งมีตัวแปรอิสระ 17 ตัวแปร และตัวแปรตาม 1 ตัวแปร คือค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์ โดยมีตัวอย่างจำนวน 2,348,460 โซน



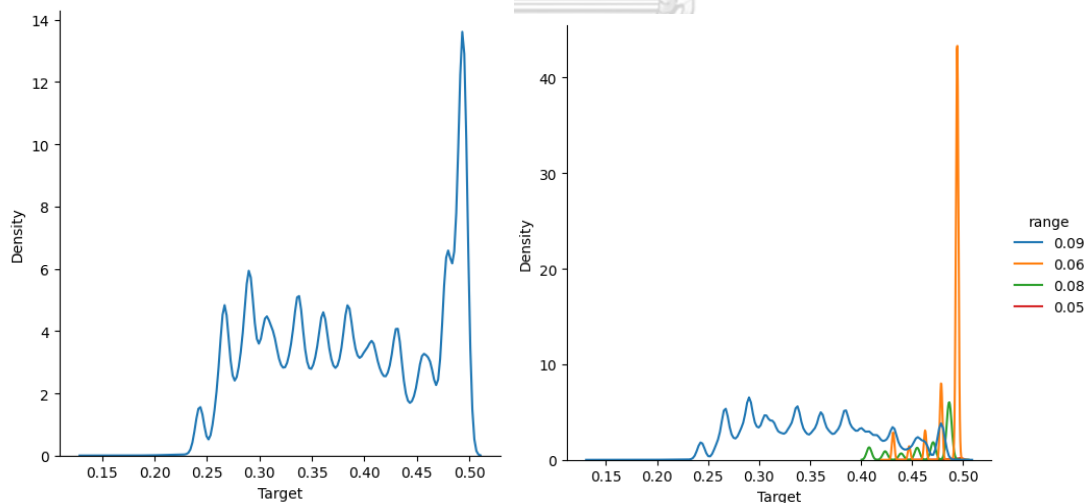
3.2.2 สำรวจ และวิเคราะห์ข้อมูลเชิงลึกของชุดข้อมูล

เราจะตรวจสอบช่วงความกว้างสำหรับการวัดค่ากระแสไฟฟ้าเขียน จากการทดสอบฮาร์ดแวร์ โดยคำนวณช่วงความกว้างสำหรับการวัดค่ากระแสไฟฟ้าเขียนจากผลต่างระหว่าง MAX และ MIN ว่าแต่ละช่วงมีความกว้างเท่าใดบ้าง และคิดเป็นอัตราร้อยละเท่าใดเมื่อเทียบกับช่วงอื่นๆ



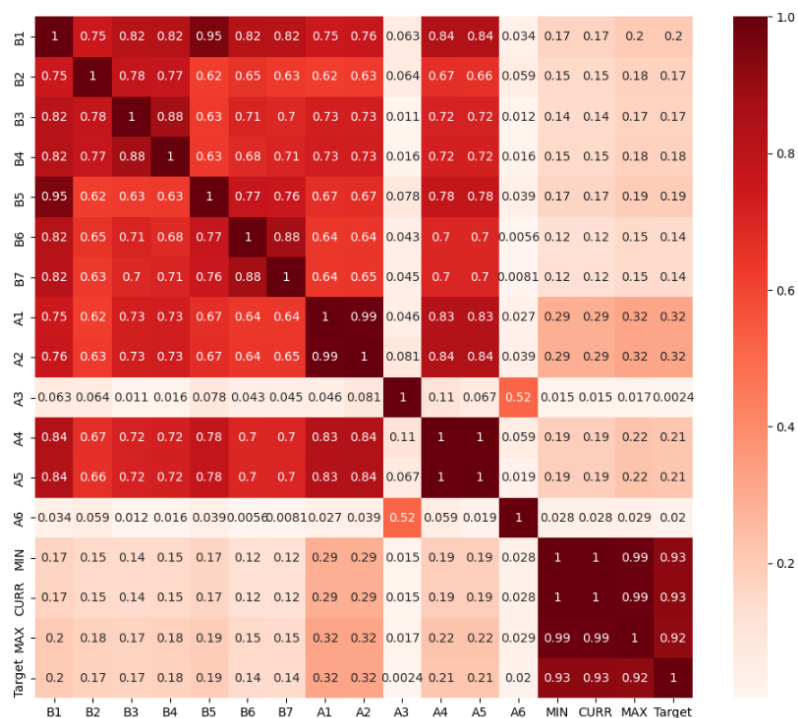
รูปที่ 11 กราฟแท่งที่แสดงอัตราส่วนของจำนวนตัวอย่างของแต่ละช่วงความกว้าง

ที่รูป 11 เราพบว่ามียู่ 4 ช่วงความกว้างต่อมาเราจะตรวจสอบการแจกแจงของตัวแปรตาม ที่แบ่งตามกลุ่มช่วงความกว้างการวัดค่า และบริเวณช่วงของดิสก์จากรูปที่ 12 พบว่าการแจกแจงของตัวแปรตามนั้น มีมากกว่า 1 การแจกแจงซ้อนทับกัน เราจึงแยกการแจกแจงตามช่วงความกว้างวัดค่า จากรูปที่ 13 พบว่าแต่ละช่วงความกว้างนั้นมีการแจกแจงในแต่ละช่วงความกว้างเป็นของตนเอง



รูปที่ 12 การแจกแจงของตัวแปรตาม , รูปที่ 13 การแจกแจงของตัวแปรตามแต่ละช่วงความกว้างวัดค่า

สุดท้ายแล้วเราจะตรวจสอบว่า แต่ละตัวแปรอิสระนั้นมีปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระโดยวิธี Pearson จากรูปที่ 14 เราพบว่ามีปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระค่อนข้างสูงระหว่าง B1-B7, A1-A2, A4-A5 และ MIN-MAX



รูปที่ 14 ความสัมพันธ์ระหว่างตัวแปรอิสระโดยวิธี Pearson

3.2.3 เตรียมการชุดข้อมูลให้พร้อมสำหรับเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

หลังจากค้นหาข้อมูลเชิงลึกของชุดข้อมูล ผู้วิจัยตัดสินใจที่จะตัดหน่วยตัวอย่างที่มีช่วงความกว้างในการวัดค่าที่ 0.05 ทิ้ง เนื่องจากช่วงความกว้างในการวัดนี้มีจำนวนน้อยจนการแจกแจงของตัวแปรตามไม่ชัดเจน เมื่อเปรียบเทียบกับช่วงความกว้างในการวัดค่า 0.09, 0.08, 0.06 แล้วจำนวนตัวอย่างของข้อมูลชุดนี้จะเหลือ 2,341,932 โชน แล้วผู้วิจัยจะทำการนำตัวแปรอิสระที่มีค่าความสัมพันธ์ระหว่างตัวแปรมากกว่าหรือเท่ากับ 0.95 ออก นอกจากตัวแปรอิสระ MIN, CURR, MAX เนื่องจากเป็นตัวแปรอิสระที่ผู้วิจัยตั้งใจให้ตัวแบบการถดถอยโดยความไม่แน่นอนเรียนรู้ขอบเขตช่วงการวัดค่ากระแสไฟฟ้าเขี่ยนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์ ซึ่งตัวแปรอิสระที่มีค่าความสัมพันธ์ระหว่างตัวแปรมากกว่าหรือเท่ากับ 0.95 ได้แก่ B5, A2, A5 ออก

นำตัวอย่างที่เหลือจากการดำเนินการข้างต้น โดยแบ่งเป็นชุดฝึก ชุดทดสอบ และชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ

ตารางที่ 2 จำนวนหน่วยตัวอย่างแต่ละสัปดาห์

week	number of samples
01	294512
02	302442
03	303659
04	309863
05	3348
06	1512
07	300218
08	301489
09	297806

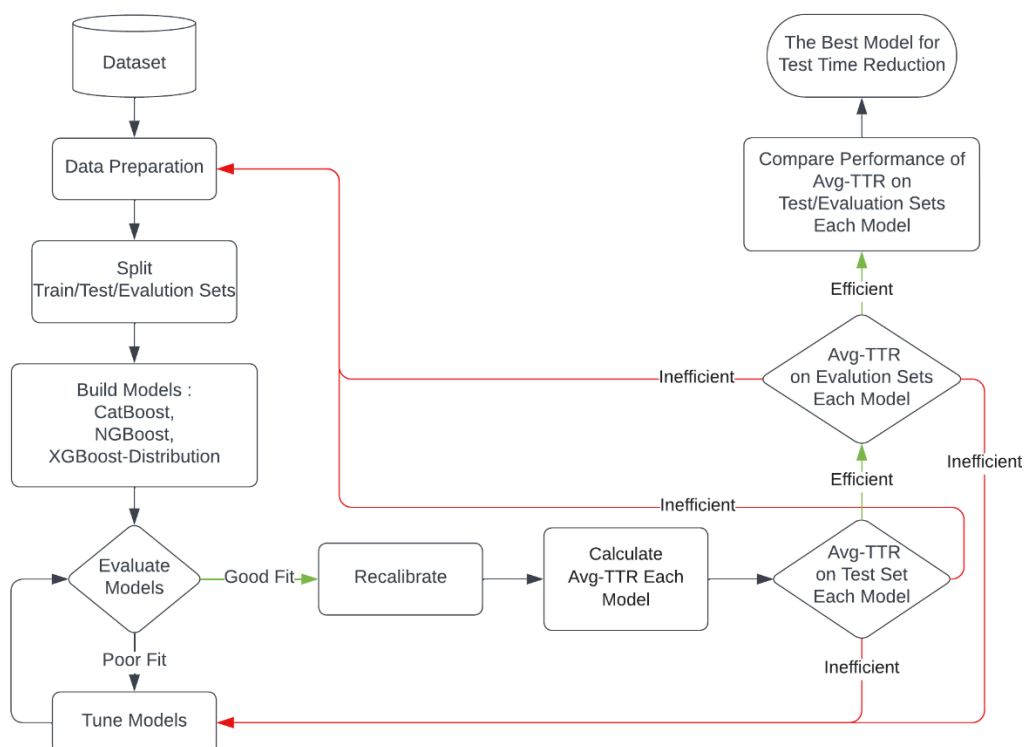
จากตารางที่ 2 โดยชุดฝึกและชุดทดสอบ จะใช้หน่วยตัวอย่างสัปดาห์ที่ทดสอบหลังประกอบฮาร์ดไดรฟ์ที่ 01, 02, 03, 04 ซึ่งมีจำนวน 1,210,476 โชน โดยแบ่งเป็นอัตราส่วนร้อยละ 75 และ 25 ตามลำดับ ซึ่งแต่ละชุดประกอบไปด้วย ตัวแปรอิสระ และตัวแปรตาม จากนั้นนำตัวแปรอิสระประเภท numeric มาปรับสเกลโดย Standard Scaler วิธี z-score และประเภท category มาเข้ารหัสตัวเลขโดย Label Encoder จากไลบรารี sklearn.preprocessing และชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ

แล้วใช้หน่วยตัวอย่างสัปดาห์ที่ทดสอบหลังประกอบฮาร์ดไดรฟ์ที่ 07, 08, 09 แบ่งเป็นสัปดาห์ละ 1 ชุด รวมมีชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ 3 ชุด จากนั้นนำตัวแปรอิสระแต่ละชุดประเภท numeric มาปรับสเกลโดย Standard Scaler และประเภท category มาเข้ารหัสตัวเลขโดย Label Encoder อ้างอิงจากชุดฝึกและชุดทดสอบ ส่วนหน่วยตัวอย่างสัปดาห์ที่ทดสอบหลังประกอบฮาร์ดไดรฟ์ที่ไม่ได้นำมาเรียนรู้ หรือทดสอบ เนื่องจากเป็นสัปดาห์ที่ทดสอบอยู่ในช่วงไม่ปกติ

3.2.4 ดำเนินฝึกการเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

ฝึกการเรียนรู้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน ด้วยวิธี CatBoost, วิธี XGB-Distribution และ วิธี NGBBoost ใช้ข้อมูลจาก ชุดฝึกและชุดทดสอบ จาก 3.2.3 โดยไฮเปอร์พารามิเตอร์จนได้ค่า *AvgTTR* ซึ่งนำค่าความไม่แน่นอนของข้อมูลที่ผ่านการ Recalibrated เพื่อนำมาเปรียบเทียบประสิทธิภาพในการลดเวลาการทดสอบกระแสไฟฟ้าเขียน กับชุดทดสอบจนได้จนได้ค่า *AvgTTR* ที่ดีที่สุดของแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน แล้วนำตัวแบบมาทดสอบการพยากรณ์เพื่อลดเวลาการทดสอบกระแสไฟฟ้าเขียน กับชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ 3 ชุด จาก 3.2.3 หากประสิทธิภาพในการลดเวลาการทดสอบกระแสไฟฟ้าเขียน ค่า *AvgTTR* ต่างจากชุดทดสอบ จะดำเนินการไฮเปอร์พารามิเตอร์ จนกว่าได้ค่า *AvgTTR* ที่ชุดทดสอบ และชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ 3 ชุด ไม่แตกต่างกัน ของแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

Comparative Study Process



รูปที่ 15 แผนผังการวิจัย

บทที่ 4 ผลการวิจัย

งานวิทยานิพนธ์นี้เป็นการศึกษาเปรียบเทียบตัวแบบจำลองการถดถอยโดยความไม่แน่นอน เพื่อลดเวลาในกระบวนการทดสอบวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดของฮาร์ดไดรฟ์ ซึ่งมีตัวแบบจำลองการถดถอยโดยความไม่แน่นอนที่ใช้การลดเวลาการทดสอบกระแสไฟฟ้าเขียนมีทั้งหมด 3 วิธี ได้แก่ NGBoost, XGBoost และCatBoost ซึ่งเกณฑ์ที่ใช้ในการพิจารณาหลักคือ ค่า *AvgTTR* และเกณฑ์ที่ใช้ในการพิจารณารองคือ MAE และRMSE โดยมีเกณฑ์ดังนี้

ถ้า ค่า *AvgTTR* ของแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอนใด มีความสามารถลดเวลาการทดสอบกระแสไฟฟ้าเขียนมากที่สุด(นาที) กว่าตัวแบบจำลองอื่นๆ แล้ว ถือว่าเป็นแบบจำลองการถดถอยโดยความไม่แน่นอนนั้น มีประสิทธิภาพในการลดเวลาการทดสอบสูงสุด

แต่ถ้า ค่า *AvgTTR* ของแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน มีความสามารถลดเวลาการทดสอบกระแสไฟฟ้าเขียนใกล้เคียงกันที่สุด(นาที) ต้องต่างกันไม่เกิน 2 นาที ต้องนำเกณฑ์ที่ใช้ในการพิจารณารอง มาพิจารณาว่า ตัวแบบใดมีค่า MAE และ RMSE น้อยที่สุด จะถือว่า เป็นแบบจำลองการถดถอยโดยความไม่แน่นอนนั้น มีประสิทธิภาพในการลดเวลาการทดสอบสูงสุด

ข้อกำหนดเบื้องต้น

1. ข้อมูลที่ใช้สำหรับการเรียนรู้ของตัวแบบจำลองการถดถอยโดยความไม่แน่นอน และทดสอบในวิทยานิพนธ์นี้ ตัวแปรอิสระที่ใช้ประกอบไปด้วย ข้อมูลประสิทธิภาพของแต่ละส่วนประกอบฮาร์ดไดรฟ์ก่อนประกอบเป็นฮาร์ดไดรฟ์ ข้อมูลประสิทธิภาพของฮาร์ดไดรฟ์หลังประกอบ และข้อมูลช่วงกระแสไฟฟ้าเขียนที่วัดได้จากการทดสอบฮาร์ดไดรฟ์หลังประกอบ
2. การแบ่งสัดส่วนชุดข้อมูลในสำหรับเรียนรู้ และสำหรับการทดสอบ เป็นร้อยละ 75 และ 25 ตามลำดับ จากชุดข้อมูลสัปดาห์ที่ 01, 02, 03, 04
3. ชุดการทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ จากชุดข้อมูลสัปดาห์ที่ 07, 08, 09 แบ่งเป็นสัปดาห์ละ 1 ชุด รวมเป็น 3 ชุด

4. ระดับความเชื่อมั่นที่นำมาเปรียบเทียบ $cl \in \{0.1, 0.11, 0.12, \dots, 0.97, 0.98, 0.99\}$
5. เกณฑ์ในการเปรียบเทียบวิธีตัวแบบจำลองการถดถอยโดยความไม่แน่นอนในวิทยานิพนธ์นี้ เปรียบความสามารถในการลดเวลาในการวัดค่ากระแสไฟฟ้าเขียนที่เหมาะสมที่สุดว่า ตัวแบบจำลองใดที่สามารถลดเวลาได้มากที่สุด
6. ในวิทยานิพนธ์นี้จะทำการปรับสเกลข้อมูลเชิงปริมาณ และแปลงข้อมูลเชิงคุณภาพ ในการแสดงดำเนินการ และผลลัพธ์ เพื่อหลีกเลี่ยงตัวอย่างข้อมูลที่ไม่สามารถเผยแพร่ได้

ข้อจำกัดของงานวิจัย

1. ไม่สามารถใช้ตัวแบบจำลองการถดถอยทั่วไป ที่ให้ผลลัพธ์แค่ค่าประมาณแบบจุด เนื่องจากมีความเสี่ยงสูงที่ค่าประมาณแบบจุดที่ได้ตัวแบบจำลองการถดถอยทั่วไป อาจจะไม่ใช้กระแสไฟฟ้าเขียนที่เหมาะสมที่สุด ซึ่งจะส่งผลกระทบต่อประสิทธิภาพการทำงานของฮาร์ดไดรฟ์
2. ไม่สามารถใช้ตัวแบบจำลองการเรียนรู้เชิงลึก (Deep learning) เนื่องจาก ฮาร์ดแวร์ของเครื่องทดสอบฮาร์ดไดรฟ์ที่ติดตั้งตัวแบบจำลองนั้น ไม่สามารถรองรับทรัพยากร และความต้องการของตัวแบบจำลองการเรียนรู้เชิงลึก ถ้า ต้องการที่จะติดตั้งตัวแบบจำลองการเรียนรู้เชิงลึกจะต้องเพิ่มประสิทธิภาพ หรือเปลี่ยนฮาร์ดแวร์ แล้วต้องเพิ่มค่าใช้จ่ายสำหรับการติดตั้งฮาร์ดแวร์ของเครื่องทดสอบ
3. ไม่สามารถเพิ่มตัวแปรอิสระนอกเหนือจากที่กำหนดได้ เนื่องจากมีค่าใช้จ่ายในการเก็บข้อมูลแต่ละการทดสอบ

4.1 ผลลัพธ์การประเมินแบบจำลองการถดถอยโดยความไม่แน่นอน ก่อนนำค่าความไม่แน่นอนของข้อมูลไปใช้คำนวณ $AvgTTR$

เราฝึกตัวแบบจำลองการถดถอยโดยความไม่แน่นอนบน CPU 80 core แล้วนำมาประเมินประสิทธิภาพของตัวแบบจำลองก่อนที่เราทำการเปรียบเทียบค่า $AvgTTR$ ของแต่ละตัวแบบจำลอง โดยเราจะนำเกณฑ์คะแนน MAE และ RMSE มาบ่งบอกประสิทธิภาพของตัวแบบจำลอง โดยจะประเมินแต่ละตัวแบบจำลองว่า เป็นตัวแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ (Good fit) ก็ต่อเมื่อ ค่าเกณฑ์คะแนนของชุดฝึก และทุกชุดตัวอย่างทดสอบ ถ้า มีค่า MAE น้อยกว่าหรือเท่ากับ 0.025 และ ค่า RSME น้อยกว่าหรือเท่ากับ 0.03 แล้ว ค่า MAE และ RSME ของชุดฝึก และทุกชุดตัวอย่างทดสอบ แต่ละชุดมีค่าแตกต่างกันไม่เกิน ± 0.0005 (วิทยานิพนธ์นี้ได้จะทำการปรับสเกลข้อมูลเชิงปริมาณ ค่า MAE และ RSME เพื่อปกป้องความลับของบริษัท) หากตัวแบบจำลองไม่มีประสิทธิภาพในการพยากรณ์ (Poor fit) จะทำการดำเนินการไฮเปอร์พารามิเตอร์ค่าที่เหมาะสมที่สุด เพื่อให้ค่า $AvgTTR$ ได้มากที่สุดสำหรับการลดเวลาทดสอบกระแสไฟฟ้าเขียนแต่ละตัวแบบจำลอง จนกว่าจะได้ตัวแบบจำลองที่มีประสิทธิภาพในการพยากรณ์

ตารางที่ 3 ผลลัพธ์เกณฑ์คะแนน MAE และ RSME ของชุดฝึก และทุกชุดตัวอย่างทดสอบ แต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน

ชุดตัวอย่างทดสอบ	เกณฑ์คะแนน	ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน		
		NGBoost	XGBoost-Distribution	CatBoost
ชุดฝึก	MAE	0.0234	0.0233	0.0233
	RMSE	0.0286	0.0284	0.0285
ชุดทดสอบ	MAE	0.0235	0.0234	0.0234
	RMSE	0.0287	0.0286	0.0286
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07	MAE	0.0231	0.0231	0.0230
	RMSE	0.0285	0.0285	0.0284
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08	MAE	0.0231	0.0230	0.0230
	RMSE	0.0284	0.0283	0.0283
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09	MAE	0.0231	0.0230	0.0230
	RMSE	0.0284	0.0284	0.0283

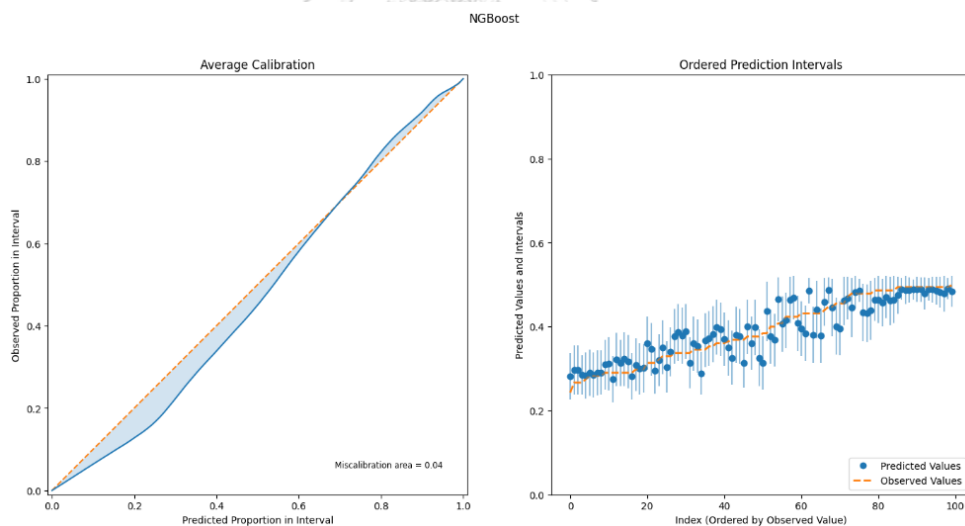
*หมายเหตุ : ในวิทยานิพนธ์นี้จะทำการปรับสเกลข้อมูลเชิงปริมาณ เพื่อปกป้องความลับของบริษัท

จากตารางที่ 3 จะแสดงผลลัพธ์เกณฑ์คะแนน MAE และ RSME ของชุดฝึก และทุกชุดตัวอย่างทดสอบแต่ละตัวแบบจำลองจะเห็นว่า ทุกตัวแบบจำลองการถดถอยโดยความไม่แน่นอน มีค่า MAE น้อยกว่า 0.025 และ ค่า RSME น้อยกว่า 0.03 ในชุดฝึก และทุกชุดตัวอย่างทดสอบ อีกทั้งแต่ละชุดมีค่าแตกต่างกันไม่เกิน ± 0.0005

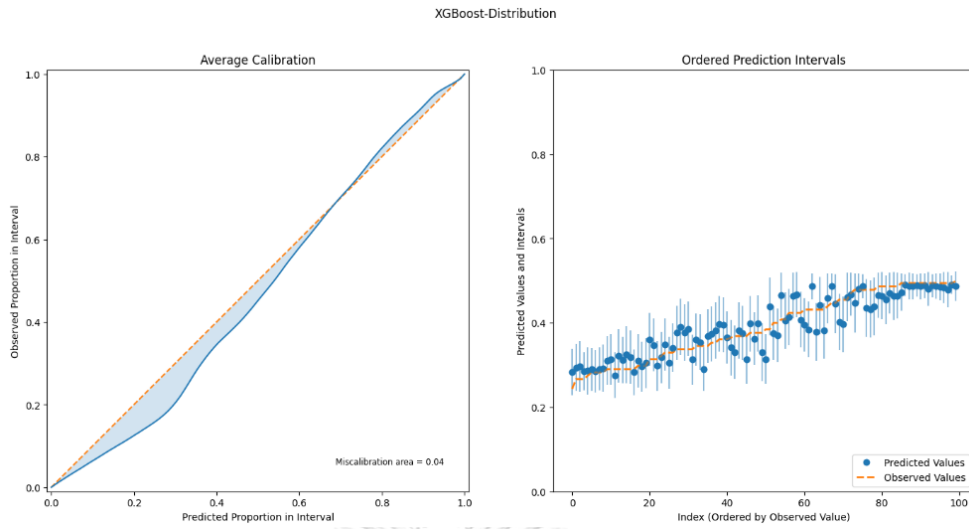
ดังนั้นสามารถกล่าวได้ว่า NGBoost, XGBoost-Distribution และ CatBoost เป็นตัวแบบจำลองที่มีประสิทธิภาพในการพยากรณ์(Good fit) ซึ่งสามารถนำค่าความไม่แน่นอนของข้อมูลจากแต่ละตัวแบบจำลองไปดำเนินการต่อไป Recalibrated ค่าความไม่แน่นอนของข้อมูล แล้วนำไปคำนวณค่า *AvgTTR* สำหรับการลดเวลาทดสอบกระแสไฟฟ้าเขียนแต่ละตัวแบบจำลอง เพื่อเปรียบเทียบความสามารถในการลดเวลาการทดสอบกระแสไฟฟ้าเขียนระหว่าง NGBoost, XGBoost-Distribution และ CatBoost

4.2 ผลลัพธ์การเปรียบเทียบความผิดพลาดของการเทียบมาตรฐาน (Miscalibration Area)

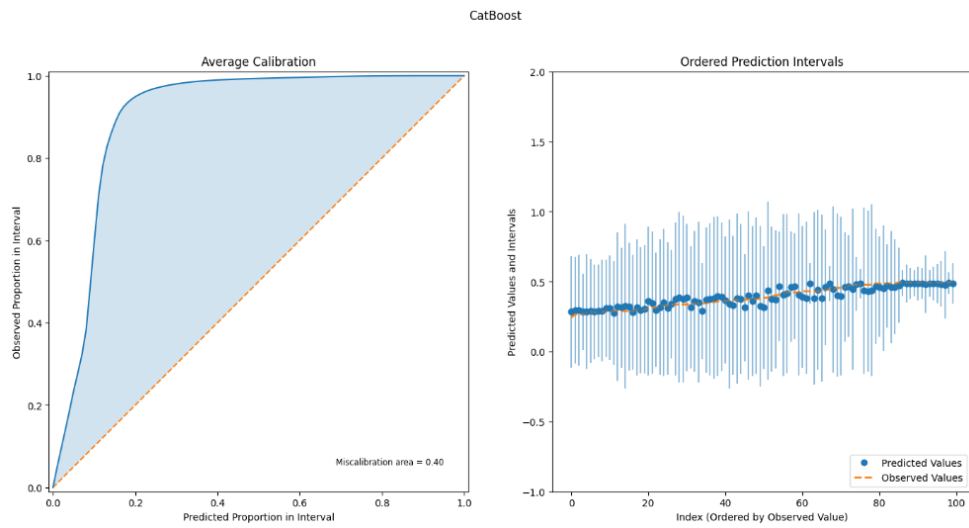
เราจะแสดงผลลัพธ์การพยากรณ์ของชุดทดสอบ แต่ละตัวแบบจำลอง ได้แก่ RMSE, MAE และ Miscalibration Area (MA) ก่อนจะทำการ Recalibrated ค่าความไม่แน่นอนของข้อมูล



รูปที่ 16 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ

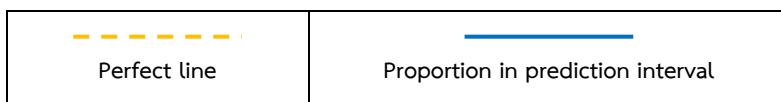


รูปที่ 17 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ

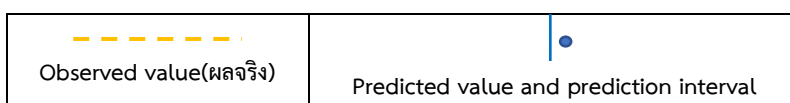


รูปที่ 18 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ

กราฟด้านซ้ายมือ



กราฟด้านขวามือ



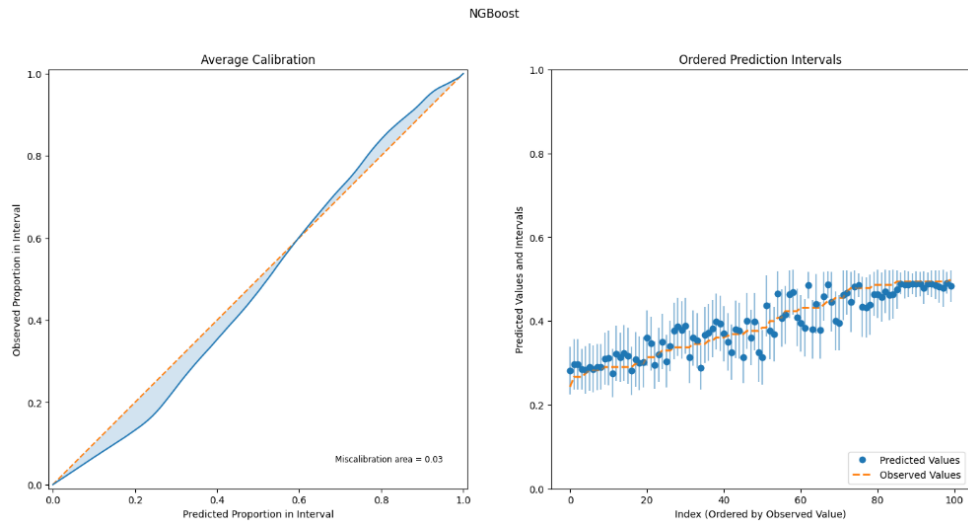
ตารางที่ 4 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ของตัวแบบจำลองความไม่แน่นอน ก่อน Recalibration

ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	MAE	RMSE	MA (Miscalibration Area)
NGBoost	0.0245	0.0286	0.0361
XGBoost-Distribution	0.0234	0.0286	0.0380
CatBoost	0.0234	0.0286	0.4032

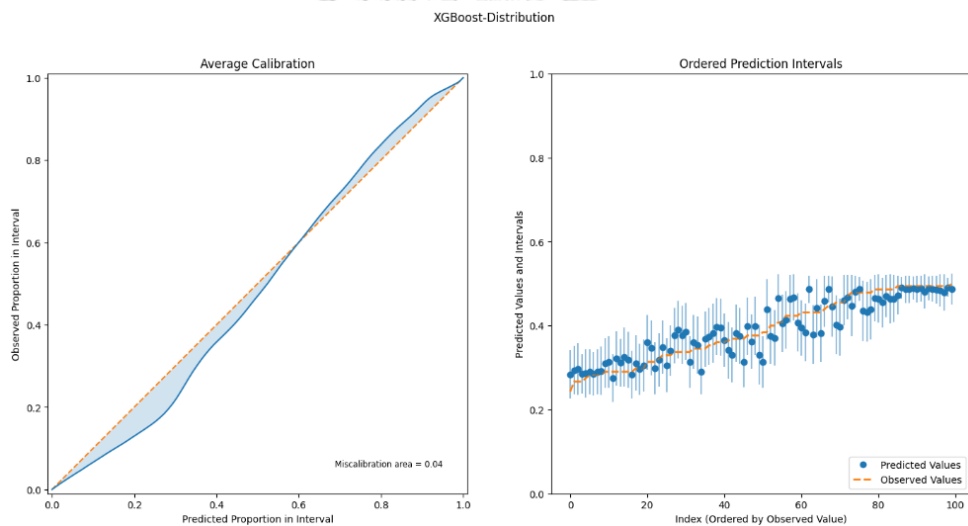
*หมายเหตุ : ในวิทยานิพนธ์นี้จะทำการปรับสเกลข้อมูลเชิงปริมาณ เพื่อปกป้องความลับของบริษัท

จากรูปที่ 16-18 และตารางที่ 4 ผลลัพธ์เกณฑ์คะแนน MAE และ RMSE แต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอนไม่แตกต่างกัน นอกจาก NGBoost ที่มีค่า MAE เท่ากับ 0.0245 ที่มีค่ามากกว่าตัวแบบจำลองอื่นเล็กน้อย อย่างไรก็ตาม CatBoost มีค่า MA เท่ากับ 0.4032 บ่งบอกถึงปัญหาที่แสดง Calibrated ที่เรียกว่า Overconfidence กล่าวคือ สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงมากกว่าสัดส่วน ณ ระดับความเชื่อมั่นจริงนั้นๆ แต่ในทางตรงกันข้าม NGBoost มีค่า MA เท่ากับ 0.0361 และ XGBoost-Distribution มีค่า MA เท่ากับ 0.0380 ซึ่งแทบจะไม่มีปัญหา หรือเรียกได้ว่าแสดง Calibrated เป็น Well-calibration กล่าวคือ สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงเท่ากับ หรือใกล้เคียงสัดส่วน ณ ระดับความเชื่อมั่นจริงนั้นๆ

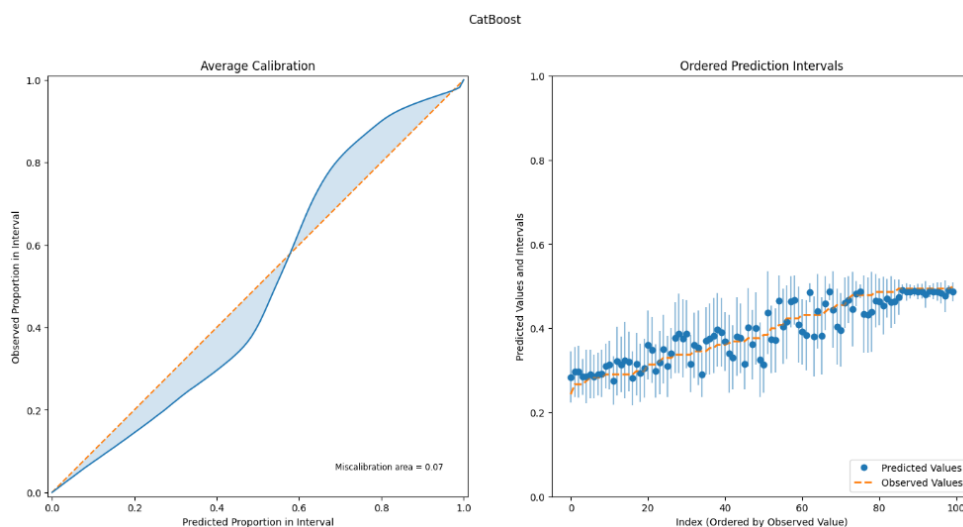
จะแสดงให้เห็นว่า XGBoost-Distribution และ NGBoost นั้นเหมาะสมกว่าสำหรับการคำนวณช่วงความเชื่อมั่นของผลทำนายมากกว่า CatBoost เนื่องจากการประมาณความไม่แน่นอนของข้อมูลที่เหมาะสมกว่า CatBoost อย่างไรก็ตาม สิ่งสำคัญคือ ต้องทำการ Recalibrated ความไม่แน่นอนของข้อมูลใหม่ให้อยู่ในระดับที่เหมาะสม โดยต้องสามารถครอบคลุมค่าที่แท้จริงตามสัดส่วนของระดับความเชื่อมั่น ก่อนนำแบบจำลองการถดถอยที่ไม่แน่นอนมาใช้เพื่อลดช่วงการวัดค่ากระแสไฟฟ้าเขียน สำหรับการลดเวลาในการทดสอบกระแสไฟฟ้าเขียน



รูปที่ 19 แสดงผลลัพธ์การพยากรณ์ของ NGBost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration



รูปที่ 20 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration



รูปที่ 21 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบ หลัง Recalibration

ตารางที่ 5 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ของตัวแบบจำลองความไม่แน่นอน หลัง Recalibration

ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน	MAE	RMSE	MA (Miscalibration Area)
NGBoost	0.0245	0.0286	0.0345
XGBoost-Distribution	0.0234	0.0286	0.0358
CatBoost	0.0234	0.0286	0.0652

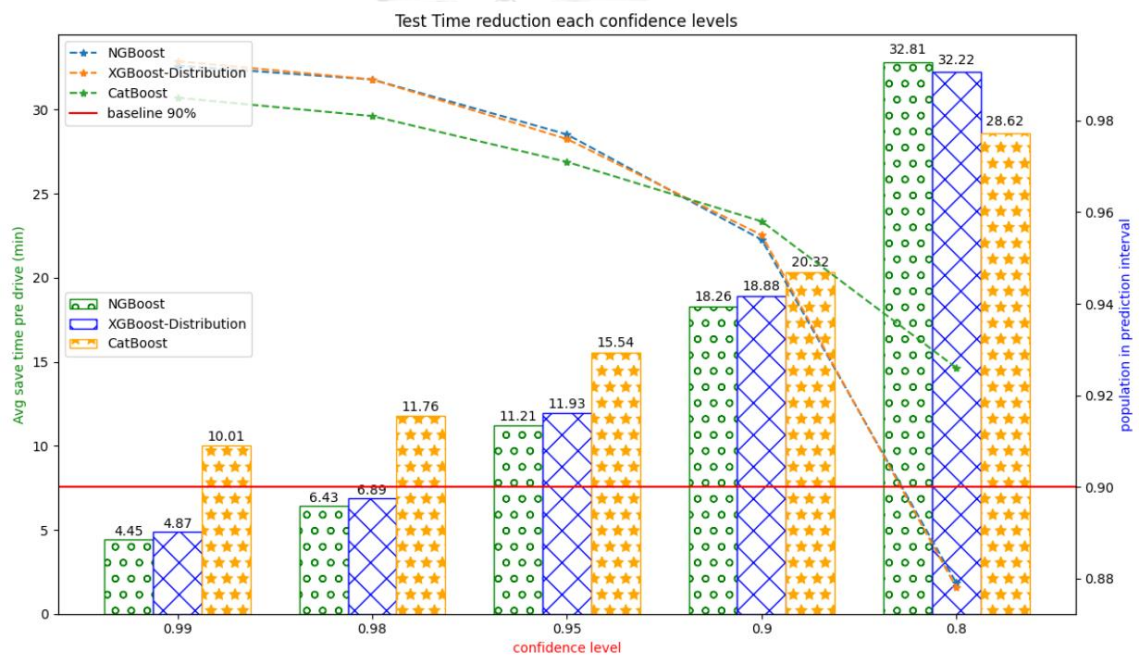
*หมายเหตุ : ในวิทยานิพนธ์นี้จะทำการปรับสเกลข้อมูลเชิงปริมาณ เพื่อปกป้องความลับของบริษัท

จากรูปที่ 19-21 และตารางที่ 5 ผลลัพธ์เกณฑ์คะแนน MAE และ RMSE ยังคงไม่เปลี่ยนแปลง หลัง Recalibrated ความไม่แน่นอนของข้อมูลที่ได้มาจากแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอน ซึ่งค่า MA ของแต่ละตัวแบบจำลอง NGBoost มีค่าเท่ากับ 0.0345 และ XGBoost-Distribution มีค่าเท่ากับ 0.0358 ลดลงจากเดิมเล็กน้อย แต่ ค่า MA ของตัวแบบจำลอง CatBoost ลดลงอย่างมากจากเดิม 0.4032 เหลือเพียง 0.0652 ซึ่งแก้ไขปัญหา Overconfidence แล้ว ดังนั้นทุกค่าความไม่แน่นอนของข้อมูลของตัวแบบจำลองที่ผ่านการ Recalibration โดยวิธีการถดถอยแบบไอโซโทนิก นั้นทำให้แสดง Calibrated เป็น Well-calibration อย่างไรก็ตามแม้ Recalibration ค่า MA ของ CatBoost ยังคงสูงกว่า XGBoost-Distribution และ NGBoost

4.3 ผลการเปรียบเทียบความสามารถในการลดเวลาการทดสอบกระแสไฟฟ้าเขียนของชุดทดสอบ แต่ระดับความเชื่อมั่น

ค่าความไม่แน่นอนของข้อมูลแต่ละตัวแบบจำลองที่ผ่านการ Recalibrated แล้วเหมาะสม
สำหรับการคำนวณช่วงความเชื่อมั่นของผลทำนาย และลดช่วงการวัดค่ากระแสไฟฟ้าเขียน โดย
สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริง ใกล้เคียงกับระดับความเชื่อมั่นนั้นๆ

ต่อมาเพื่อเปรียบเทียบความสามารถของแต่ละตัวแบบจำลองการถดถอยความไม่แน่นอน
การลดเวลาทดสอบเฉลี่ยถูกคำนวณสำหรับแต่ละระดับความเชื่อมั่นโดยช่วงความเชื่อมั่นของผล
ทำนายต้องครอบคลุมอย่างน้อย 90% ของชุดทดสอบ เพื่อแสดงว่าช่วงความเชื่อมั่นของผลทำนายมี
ประสิทธิภาพ และสามารถใช้ประเมินผลการทำงานของแต่ละตัวแบบจำลองได้



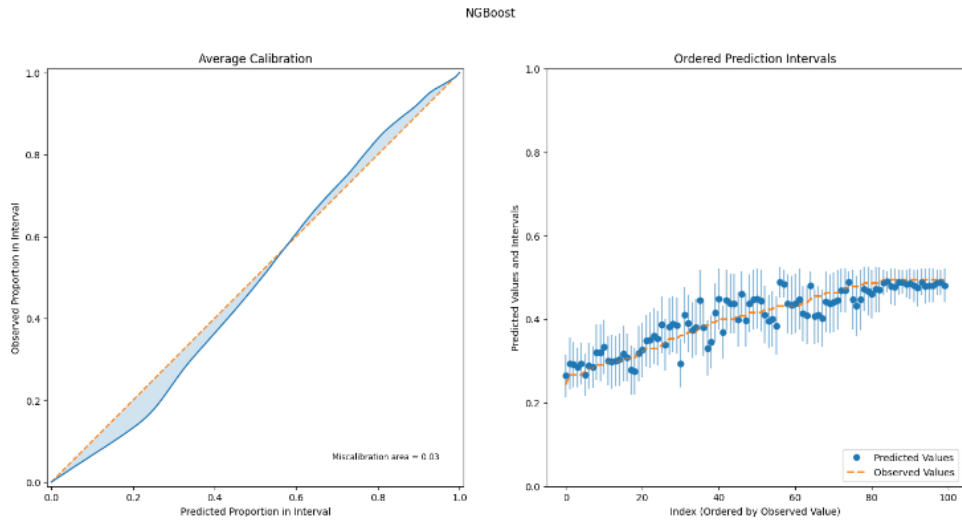
รูปที่ 22 แสดงการลดเวลาการทดสอบแต่ละตัวแบบจำลอง แต่ระดับความเชื่อมั่น

ตารางที่ 6 ผลลัพธ์การลดเวลาการทดสอบ แต่ละระดับความเชื่อมั่น

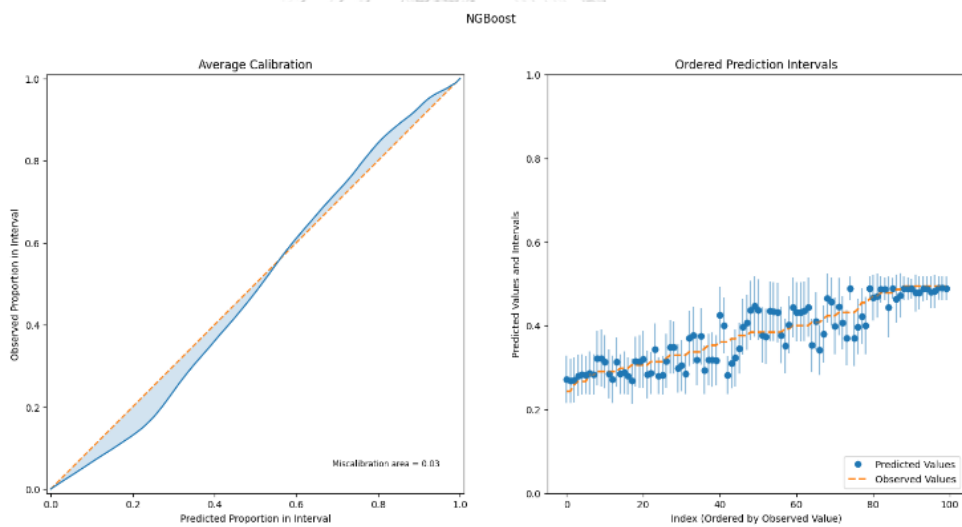
เกณฑ์คะแนน	ระดับความเชื่อมั่น	ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน		
		NGBoost	XGBoost-Distribution	CatBoost
การลดเวลาในการทดสอบ(นาที)	0.99	4.45	4.87	10.01
	0.98	6.43	6.89	11.76
	0.95	11.21	11.93	15.54
	0.90	18.26	18.88	20.32
	0.80	32.81	32.22	28.62
สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริง	0.99	0.992	0.993	0.985
	0.98	0.989	0.989	0.981
	0.95	0.977	0.976	0.971
	0.90	0.954	0.955	0.958
	0.80	0.879	0.878	0.926

ภาพที่ 22 และตารางที่ 6 แสดงให้เห็นว่า CatBoost ดีกว่า NGBoost และ XGBoostDistribution ในการลดเวลาทดสอบที่ระดับความเชื่อมั่น 0.99, 0.98, 0.95 และ 0.90 แต่ลดเวลาได้ต่ำสุดที่ระดับความมั่นใจ 0.80 อย่างไรก็ตาม ช่วงความเชื่อมั่นของผลทำนายของ CatBoost เท่านั้นที่ครอบคลุมผลจริงของชุดทดสอบอย่างน้อย 90% เท่านั้น การเปรียบเทียบความสามารถในการลดเวลาทดสอบที่แต่ละระดับความเชื่อมั่น อาจไม่เหมาะสมเนื่องจากสัดส่วนของช่วงความเชื่อมั่นของผลทำนายไม่สามารถเปรียบเทียบได้ เนื่องจากไม่เท่ากัน หรือใกล้เคียงกัน ดังนั้นเราจึงวิธีการเปรียบเทียบตัวแบบจำลอง โดยใช้ค่า *AvgTTR* ที่เปรียบเทียบการลดเวลาทดสอบที่ระดับความมั่นใจที่ต่ำสุด ที่ทำให้สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงของชุดทดสอบอย่างน้อย 90% ของแต่ละตัวแบบจำลองการถดถอยความไม่แน่นอน

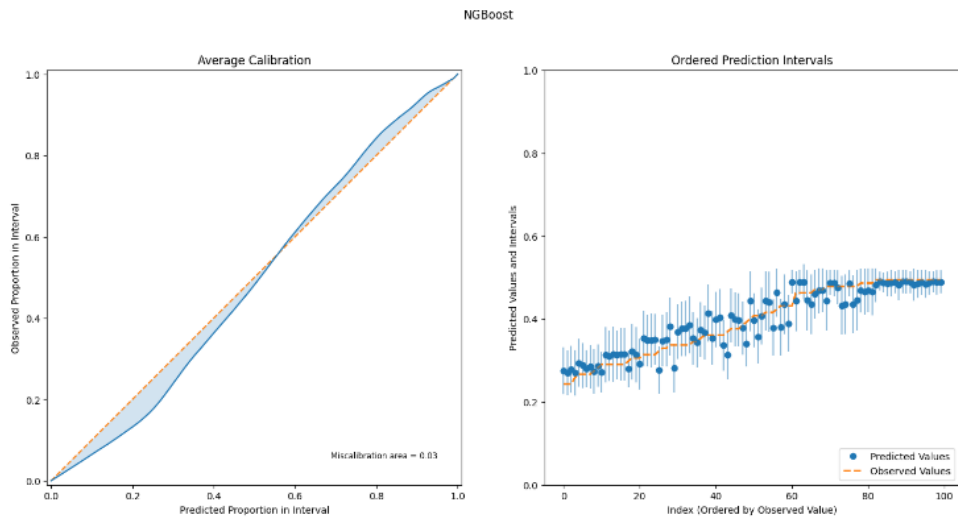
4.4 ผลการเปรียบเทียบความสามารถในการลดเวลาการทดสอบกระแสไฟฟ้าเขียนของชุดทดสอบ และชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบ



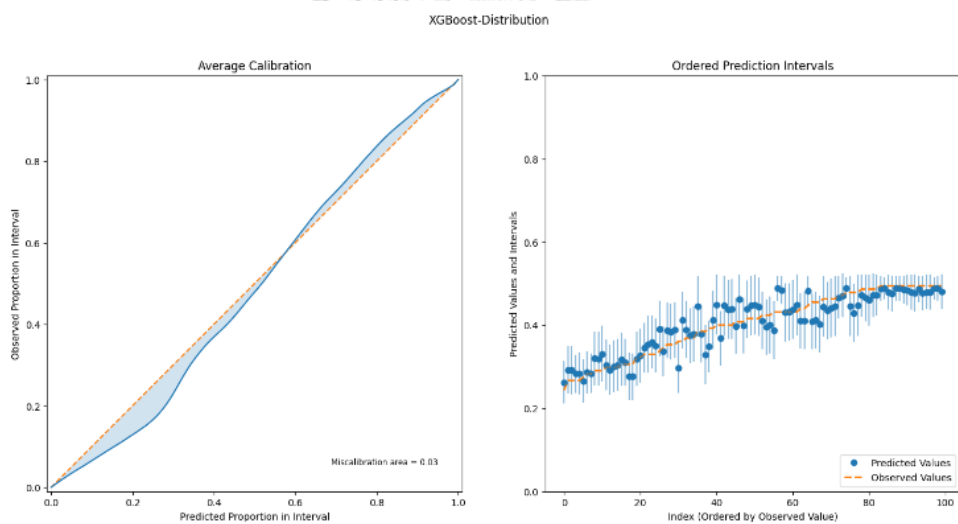
รูปที่ 23 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration



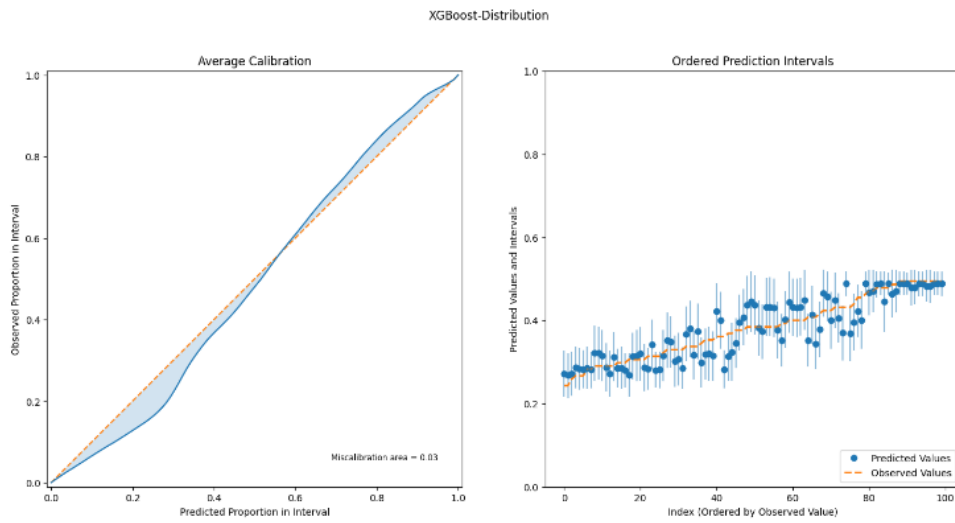
รูปที่ 24 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration



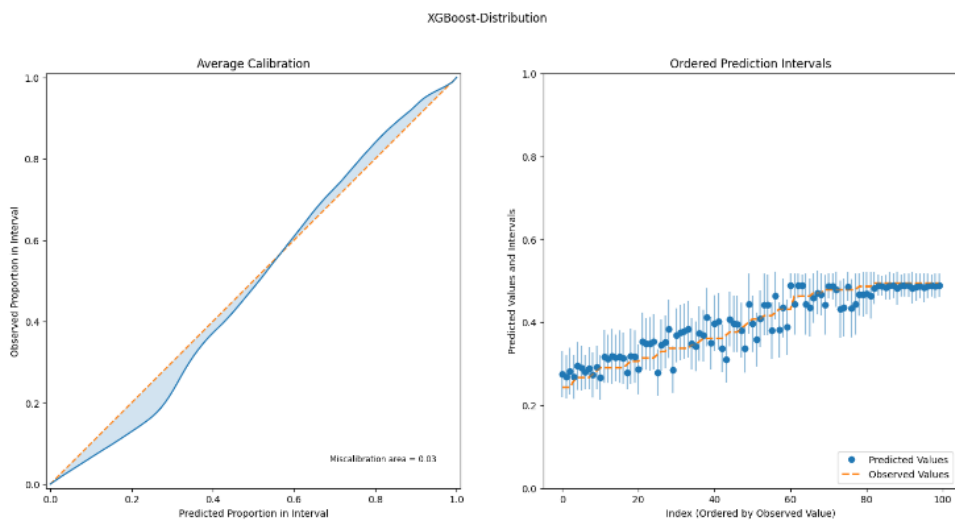
รูปที่ 25 แสดงผลลัพธ์การพยากรณ์ของ NGBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration



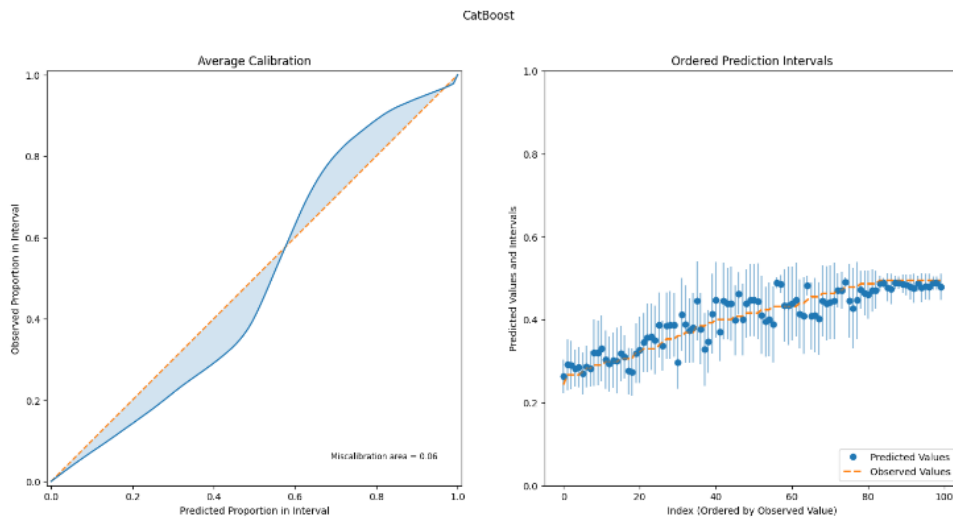
รูปที่ 26 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration



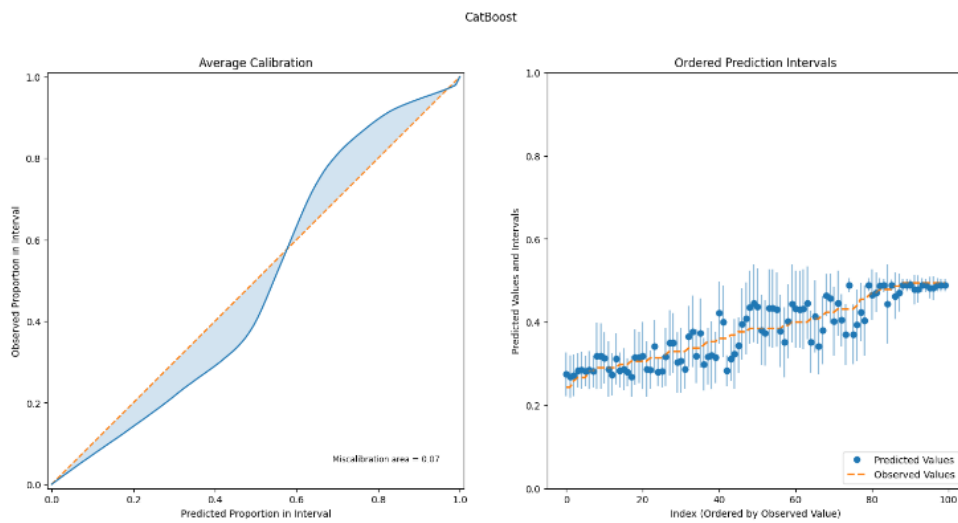
รูปที่ 27 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration



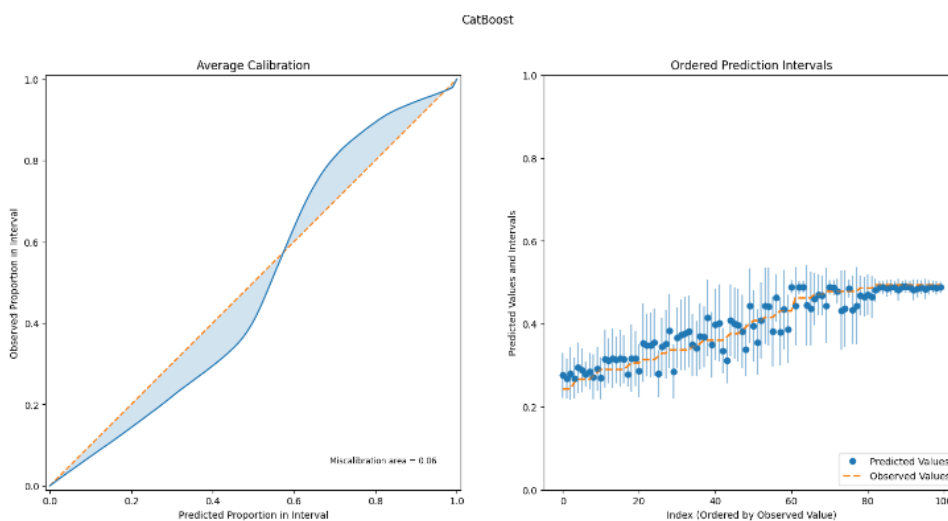
รูปที่ 28 แสดงผลลัพธ์การพยากรณ์ของ XGBoost-Distribution โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration



รูปที่ 29 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07 หลัง Recalibration



รูปที่ 30 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08 หลัง Recalibration



รูปที่ 31 แสดงผลลัพธ์การพยากรณ์ของ CatBoost โดยแสดงกราฟ Calibration และ กราฟระหว่างผลจริง และผลพยากรณ์ ของชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09 หลัง Recalibration

ตารางที่ 7 ผลลัพธ์เกณฑ์คะแนนผลจริงที่ได้จากการพยากรณ์ และการลดเวลาการทดสอบของตัวแบบจำลองความไม่แน่นอน แต่ละชุดตัวอย่างทดสอบ

ชุดตัวอย่างทดสอบ	เกณฑ์คะแนน	ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน		
		NGBoost	XGBoost-Distribution	CatBoost
ชุดทดสอบ	MAE	0.0235	0.0234	0.0234
	RMSE	0.0287	0.0286	0.0286
	MA	0.0345	0.0358	0.0652
	<i>AvgTTR</i>	27.97	28.35	32.13
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07	MAE	0.0231	0.0231	0.0230
	RMSE	0.0285	0.0285	0.0284
	MA	0.0316	0.0329	0.0638
	<i>AvgTTR</i>	28.46	28.95	33.6
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08	MAE	0.0231	0.0230	0.0230
	RMSE	0.0284	0.0283	0.0283
	MA	0.0337	0.0347	0.0663
	<i>AvgTTR</i>	29.56	29.91	34.19
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09	MAE	0.0231	0.0230	0.0230
	RMSE	0.0284	0.0284	0.0283
	MA	0.0332	0.0336	0.0647
	<i>AvgTTR</i>	29.42	29.76	33.14

*หมายเหตุ : ในวิทยานิพนธ์นี้จะทำการปรับสเกลข้อมูลเชิงปริมาณ เพื่อปกป้องความลับของบริษัท

ในรูปที่ 19-21, 23-31 และตารางที่ 7 แสดงการเปรียบเทียบ ค่าเกณฑ์คะแนน *AvgTTR*, MAE และ RMSE ของแต่ละตัวแบบจำลองการถดถอยโดยความไม่แน่นอนของชุดทดสอบ และชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07, 08, 09 จะเห็นได้ชัดว่าเกณฑ์คะแนน MAE และ RMSE ไม่แตกต่างกันอย่างมีนัยสำคัญสำหรับแต่ละตัวแบบจำลองในชุดตัวอย่างทดสอบ หลังจากที่ Recalibrated ความไม่แน่นอนของข้อมูลสำหรับแต่ละตัวแบบจำลอง ซึ่ง NGBBoost มีค่า MA น้อยที่สุด รองลงมาคือ XGBoost-Distribution แต่ CatBoost มีค่า MA ความแตกต่างเกือบสองเท่ากับตัวแบบจำลองอื่น โดยรวมแล้ว CatBoost มีความสามารถในการลดเวลาทดสอบมากที่สุด เนื่องจากมีค่า *AvgTTR* มากที่สุดในทุกชุดตัวอย่างทดสอบ เมื่อเปรียบเทียบกับตัวแบบจำลองอื่นในการศึกษาวิทยานิพนธ์นี้



บทที่ 5 สรุปผลการดำเนินงาน และข้อเสนอแนะ

5.1 สรุปและอภิปรายผลการดำเนินงาน

สรุปของการศึกษานี้มีเป้าหมาย เพื่อใช้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน (Uncertainty Regression models) เพื่อลดเวลาทดสอบของการทดสอบกระแสไฟฟ้าเขียนที่ใช้ในกระบวนการผลิตฮาร์ดดิสก์ ซึ่งถ้าใช้ตัวแบบจำลองการถดถอยทั่วไปอาจมีปัญหาด้านประสิทธิภาพการใช้วิธี Adaptive Test Condition ที่ใช้ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน เพื่อลดช่วงการวัดของการทดสอบกระแสไฟฟ้าเขียน จากผลลัพธ์ในบทที่ 4 ซึ่ง NGBoost, XGBoost-Distribution และ CatBoost เป็นตัวแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ (Good fit) จากบทที่ 4.1

ตารางที่ 8 ผลลัพธ์เกณฑ์คะแนน Miscalibration Area ก่อนทำ และหลังทำ Recalibration ในบทที่ 4.2

Recalibration	ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน		
	NGBoost	XGBoost-Distribution	CatBoost
ก่อนทำ	0.0361	0.0380	0.4032
หลังทำ	0.0345	0.0358	0.0652

จากตารางที่ 8 และผลลัพธ์ในบทที่ 4.2 จะแสดงให้เห็นว่าค่าความไม่แน่นอนของข้อมูลของ CatBoost ก่อน Recalibrated นั้นมีปัญหา Calibrated เป็น Overconfidence แล้ว หลังจาก Recalibrated ค่าความไม่แน่นอนของข้อมูลของแต่ละตัวแบบจำลอง Calibrated เป็น Well-calibration ดังนั้นก่อนจะนำค่าความไม่แน่นอนของข้อมูลที่ได้จากตัวแบบจำลองไปใช้คำนวณช่วงเชื่อมั่นของผลทำนาย ควร Recalibration ค่าความไม่แน่นอนของข้อมูลก่อนคำนวณช่วงความเชื่อมั่นของผลทำนาย เพื่อสัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงเท่ากับ หรือใกล้เคียง สัดส่วน ณ ระดับความเชื่อมั่นจริงนั้นๆ

จากรูปที่ 22 และตารางที่ 6 ในบทที่ 4.3 จะแสดงให้เห็นว่าถึงแม้ว่า ค่าความไม่แน่นอนของข้อมูลผ่าน Recalibration แล้ว สัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงใกล้เคียง สัดส่วน ณ ระดับความเชื่อมั่นจริงนั้นๆ แต่ ค่าสัดส่วนของช่วงความเชื่อมั่นของผลทำนายแต่ละตัวแบบครอบคลุมผลจริงมากกว่าระดับความเชื่อมั่นที่ 0.9 กล่าวคือ เรายังสามารถระดับความเชื่อมั่นลงได้อีก เนื่องจากเราต้องการค่าสัดส่วนของช่วงความเชื่อมั่นของผลทำนายแต่ละตัวแบบครอบคลุมผลจริงอย่างน้อย 0.9 เพื่อเพิ่มประสิทธิภาพการลดเวลาการทดสอบ และเทียบประสิทธิภาพการลด

เวลาการทดสอบที่มีค่าสัดส่วนของช่วงความเชื่อมั่นของผลทำนายแต่ละตัวแบบครอบคลุมผลจริงอย่างน้อย 0.9 ของระดับความเชื่อมั่นของแต่ละตัวแบบ

ตารางที่ 9 ผลลัพธ์เกณฑ์การลดเวลาการทดสอบของตัวแบบจำลองความไม่แน่นอน แต่ละชุดตัวอย่างทดสอบ ในบทที่ 4.4

ชุดตัวอย่างทดสอบ	เกณฑ์คะแนน	ตัวแบบจำลองการถดถอยโดยความไม่แน่นอน		
		NGBoost	XGBoost-Distribution	CatBoost
ชุดสอบ	<i>AvgTTR</i>	27.97	28.35	32.13
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 07		28.46	28.95	33.6
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 08		29.56	29.91	34.19
ชุดทดสอบที่ไม่เกี่ยวข้องกับการพัฒนาตัวแบบสัปดาห์ที่ 09		29.42	29.76	33.14

จากตารางที่ 9 แสดงให้เห็นว่าตัวแบบจำลองการถดถอยโดยค่าความไม่แน่นอน CatBoost มีผลลัพธ์การลดเวลาการทดสอบมากที่สุดในทุกชุดตัวอย่างการทดสอบ ซึ่งเป็นตัวแบบที่เหมาะสมสำหรับ Adaptive test condition ที่มีค่าสัดส่วนของช่วงความเชื่อมั่นของผลทำนายครอบคลุมผลจริงอย่างน้อย 0.9 ซึ่งต่างจากตัวแบบ NGBoost และ XGBoost-Distribution ที่มีผลลัพธ์การลดเวลาการทดสอบใกล้เคียงกันทุกในชุดตัวอย่างการทดสอบ

5.2 ข้อเสนอแนะ

ในวิทยานิพนธ์นี้ศึกษาในกรณีของการทดสอบวัดค่ากระแสไฟฟ้าเขียนเหมาะสมที่สุดของฮาร์ดไดรฟ์ บนวัสดุประเภท A เท่านั้น หากเป็นวัสดุชนิดอื่น หรือการวัดค่าการทดสอบวัดค่ากระแสไฟฟ้าอื่น อาจมีการแจกแจง และผลลัพธ์ความสามารถในการลดเวลาการทดสอบต่างกัน อย่างไรก็ตาม สำหรับตัวแบบจำลองการถดถอยโดยความไม่แน่นอนที่ใช้ในวิทยานิพนธ์นี้ ที่มีข้อจำกัดในการศึกษาเปรียบเทียบ การเรียนรู้แบบเครื่องจักร (Machine learning) เท่านั้น หากใช้แบบจำลองการเรียนรู้แบบเชิงลึก (Deep learning) ที่สามารถประมาณค่าความไม่แน่นอน หรือการเพิ่มตัวแปรอิสระเกินจากที่ศึกษาในงานวิทยานิพนธ์นี้สำหรับการเรียนรู้ของตัวแบบจำลอง อาจจะเพิ่มประสิทธิภาพการลดเวลาในการทดสอบปัจจุบันได้ ในอนาคตอาจเป็นไปได้ว่าจะมีการลดเวลาในการทดสอบปัจจุบันโดยการประมาณค่าแบบจุด หรือประมาณค่าความไม่แน่นอนได้แม่นยำมากขึ้น

นอกจากนั้นสามารถนำแนวทางการศึกษาวิทยานิพนธ์นี้ ไปใช้กับการลดเวลาทดสอบอุปกรณ์อิเล็กทรอนิกส์ หรือผลิตภัณฑ์อื่นๆ ที่ต้องการการลดช่วงการวัดค่าที่แม่นยำในขณะทดสอบ นอกจากนี้ วิธีการนี้ยังสามารถพัฒนาต่อไปเพื่อระบุจุด หรือชุดตัวอย่างที่มีความเสี่ยงสูงมากขึ้นโดยใช้ค่าความไม่แน่นอน ซึ่งจะเพิ่มประสิทธิภาพของกระบวนการทดสอบได้อีกด้วย



บรรณานุกรม

1. Pliamsup Pornchai , S.P.a.S.K., *Effect of write current to performance of perpendicular magnetic recording system*, in *Proceedings of 48th Kasetsart University Annual Conference: Architecture and Engineering*. 2010. p. 302-309.
2. Guo, C., T. Lim, and T. Chua, *Methods and devices for reducing device test time*. 2019, Google Patents.
3. ITRS, b., *The International Technology Roadmap for Semiconductors 2.0(ITRS)*, in *Semiconductor Industry Association*. 2019. p. 21-31.
4. Al Mamun, A., GuoXiao Guo, and Chao Bi, *Hard disk drive: mechatronics and control*. CRC press. 2017.
5. Kovintavewat, P., *Signal Processing for Digital Data Storage Volume I: Basic Read-Write Channels*, National Electronics and Computer Technology Center. 2007.
6. Alpaydin, E., *Machine learning*. 2021: Mit Press.
7. Polikar, R., *Ensemble learning*. Ensemble machine learning: Methods and applications, 2012: p. 1-34.
8. Malinin, A., L. Prokhorenkova, and A. Ustimenko, *Uncertainty in gradient boosting via ensembles*. arXiv preprint arXiv:2006.10562, 2020.
9. Duan, T., et al. *Ngboost: Natural gradient boosting for probabilistic prediction*. in *International Conference on Machine Learning*. 2020. PMLR.
10. Christian, D. *xgboost-distribution*. 2020 [cited 2023 24/2]; Available from: <https://xgboost-distribution.readthedocs.io/en/latest/readme.html>.
11. *CatBoost*. [cited 2023 24/2]; Available from: <https://catboost.ai/en/docs/>.
12. Kuleshov, V., N. Fenner, and S. Ermon. *Accurate uncertainties for deep learning using calibrated regression*. in *International conference on machine learning*. 2018. PMLR.
13. Madge, R., et al. *In search of the optimum test set-adaptive test methods for maximum defect coverage and lowest test cost*. in *2004 International Conferce on Test*. 2004. IEEE.
14. Benner, S. and O. Boroffice. *Optimal production test times through adaptive test programming*. in *Proceedings International Test Conference 2001 (Cat. No. 01CH37260)*. 2001. IEEE.

15. Song, T., et al., *Pattern reorder for test cost reduction through improved SVMRANK algorithm*. IEEE Access, 2020. **8**: p. 147965-147972.
16. Chen, M. and A. Orailoglu. *Test cost minimization through adaptive test development*. in *2008 IEEE International Conference on Computer Design*. 2008. IEEE.
17. Yeh, J.-C., et al. *A systematic approach to reducing semiconductor memory test time in mass production*. in *2005 IEEE International Workshop on Memory Technology, Design, and Testing (MTDT'05)*. 2005. IEEE.
18. Letchumanan, S., et al. *Adaptive test method on production system-level testing (SLT) to optimize test cost, resources and defect parts per million (DPPM)*. in *2018 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*. 2018. IEEE.
19. Jone, W.-B., et al., *An adaptive path selection method for delay testing*. IEEE Transactions on Instrumentation and Measurement, 2001. **50**(5): p. 1109-1118.
20. Wu, W.-J., C.Y. Tang, and M. Lin. *Methods for memory test time reduction*. in *IEEE International Workshop on Memory Technology, Design and Testing*. 1996. IEEE.
21. Suresh, C.K., S. Ozev, and O. Sinanoglu, *Adaptive Reduction of the Frequency Search Space for Multi- V_{dd} Digital Circuits Using Variation Sensitive Ring Oscillators*. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2016. **36**(6): p. 1043-1053.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล	ภัทรดิศ ดำรงค์ศักดิ์
วัน เดือน ปี เกิด	1 ธันวาคม 2541
สถานที่เกิด	โรงพยาบาลมหาราช นครราชสีมา
ที่อยู่ปัจจุบัน	202/11 หมู่ 1 ซอย มีสุข6 ตำบล ป่าตาล อำเภอ เมืองลพบุรี จังหวัด ลพบุรี 15000



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY