



การวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตราการขุดเจาะ
ปีโตรเลียม

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR DRILLING
RATE PREDICTION

ญาณเดช ภาควัสดุฐานา

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

การวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตราการขาดเจาะ
ปิโตรเลียม



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR DRILLING
RATE PREDICTION



A Master's Project Submitted in Partial Fulfillment of the Requirements
for the Degree of MASTER OF SCIENCE
(Data Science)

Faculty of Science, Srinakharinwirot University

2023

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตราการใช้ยาเสพติดรายปี
ของ
ญาณเดช ภาควิชาสถาปัตยกรรมศาสตร์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก
(อาจารย์ ดร.ศุภร คนธมักดี)

..... ประธาน
(รองศาสตราจารย์ ดร.สุพัฒนา เขื่อนทิวเกียรติ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรพร เหล่าหะเกียรติ)

ชื่อเรื่อง	การวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตรา การขุดเจาะปิโตรเลียม
ผู้วิจัย	ญาณเดช ภควัฒฐาปนา
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	อาจารย์ ดร. ศุภร คนธภักดิ์

การดำเนินงานการขุดเจาะปิโตรเลียม ปัจจัยหนึ่งที่มีผลโดยตรงต่อการขุดเจาะ คือ ค่าอัตราการ
เจาะ (Rate of Penetration, ROP) การทำนายค่าอัตราการเจาะที่ถูกต้องเป็นปัจจัยสำคัญที่จะส่งผลต่อการ
ควบคุมค่าใช้จ่ายในการดำเนินงานการขุดเจาะ งานวิจัยนี้เป็นการศึกษาการเรียนรู้ของเครื่องในรูปแบบจำลอง
การเรียนรู้แบบรวมกลุ่ม เช่น Random Forest, Gradient Boosting, Extreme Gradient Boosting, Adaptive
Boosting และ K-Nearest Neighbor ข้อมูลที่ใช้ในการทำนายค่าอัตราการเจาะนำมาจากบริษัท Equinor
บริเวณตอนกลางทะเลเหนือ ประเทศนอร์เวย์ โดยชุดข้อมูลชื่อว่า Volve การวิจัยนี้เป็นกาจำลองเหมือน
สถานการณ์ของความเป็นจริง โดยเฉพาะในกรณีที่มีข้อจำกัดทางด้านข้อมูล ผู้วิจัยใช้ลำดับของข้อมูล
(Sequential data) หรือข้อมูลความลึกที่ต่อเนื่องกัน (Depth-dependent data) ทำการฝึกฝนและทดสอบรูป
แบบจำลองการเรียนรู้ของเครื่องแบบต่างๆ โดยการศึกษาด้วยเทคนิค การเรียนรู้อย่างต่อเนื่อง (Continuous
Learning) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Learning) และวัดประสิทธิภาพของรูป
แบบจำลองด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) ผลของการวิจัยนี้ให้แบบจำลองการ
เรียนรู้แบบรวมกลุ่มที่เป็น Extreme Gradient Boosting (XGB) ด้วยเทคนิคการเรียนรู้แบบหน้าต่างบานเลื่อน
(Sliding Window Learning) ในการทำนายค่าอัตราการเจาะที่มีประสิทธิภาพดีที่สุด อย่างไรก็ตาม ความ
แตกต่างของขนาดข้อมูลของการเรียนรู้แบบหน้าต่างบานเลื่อนยังให้ผลการศึกษขนาดของข้อมูลที่เหมาะสม
ที่สุด

คำสำคัญ : อัตราการเจาะ, การดำเนินงานการเจาะ, รูปแบบการเรียนรู้ของเครื่อง, แบบจำลองการเรียนรู้แบบ
รวมกลุ่ม, การเรียนรู้อย่างต่อเนื่อง, การเรียนรู้แบบหน้าต่างบานเลื่อน

Title	COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR DRILLING RATE PREDICTION
Author	YANADADE PAKAWATTHAPANA
Degree	MASTER OF SCIENCE
Academic Year	2023
Thesis Advisor	Subhorn Khonthapagdee , Ph.D.

In drilling operations, the Rate of Penetration directly impacts drilling efficiency, and accurately predicting if this rate is crucial as it affects the cost-effectiveness of the process. In this study, ensemble machine learning models such as Random Forest, Gradient Boosting, Extreme Gradient Boosting, Adaptive Boosting, and K-Nearest Neighbor were used to predict ROP using a Volve dataset from the Norwegian North Sea and provided by Equinor Company. In order to simulate real-world scenarios, especially in cases with limited amounts of data, we utilized sequential data or depth-dependent data to trained and tested machine learning models using a sequential split, such as continuous learning and sliding window split techniques. The performances of the models were evaluated by calculating mean absolute error. The results of this study showed that the ensemble machine learning models, especially Extreme Gradient Boosting with sliding window split technique outperformed in predicting the Rate of Penetration. Moreover, the various sizes of the sliding window were studied to find the optimal size.

Keyword : Rate of Penetration, Drilling Operations, Machine Learning Models, Ensemble Models, Continuous Learning, Sliding Window

กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดีจากการสนับสนุน ข้อมูลทางวิชาการ ความช่วยเหลือ คำแนะนำ ตลอดจนแนวทางในการทำวิจัยและจัดทำสารนิพนธ์ของ อ.ดร.ศุภร คนธภักดี อาจารย์ที่ปรึกษา และ คณาจารย์ทุกท่านในหลักสูตรวิทยาการข้อมูล ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ รวมถึงคำแนะนำและข้อเสนอแนะที่เป็นประโยชน์ในการปรับปรุงสารนิพนธ์ให้ดียิ่งขึ้นของคณะกรรมการสอบสารนิพนธ์ การสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัยของนิสิตบัณฑิตศึกษา ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

สุดท้ายนี้ขอขอบพระคุณครอบครัวของผู้วิจัยที่เป็นกำลังใจจนสำเร็จการศึกษา และขอบคุณเพื่อนๆในสาขาวิชาที่คอยให้ความช่วยเหลือ และให้คำแนะนำด้วยดีเสมอมา

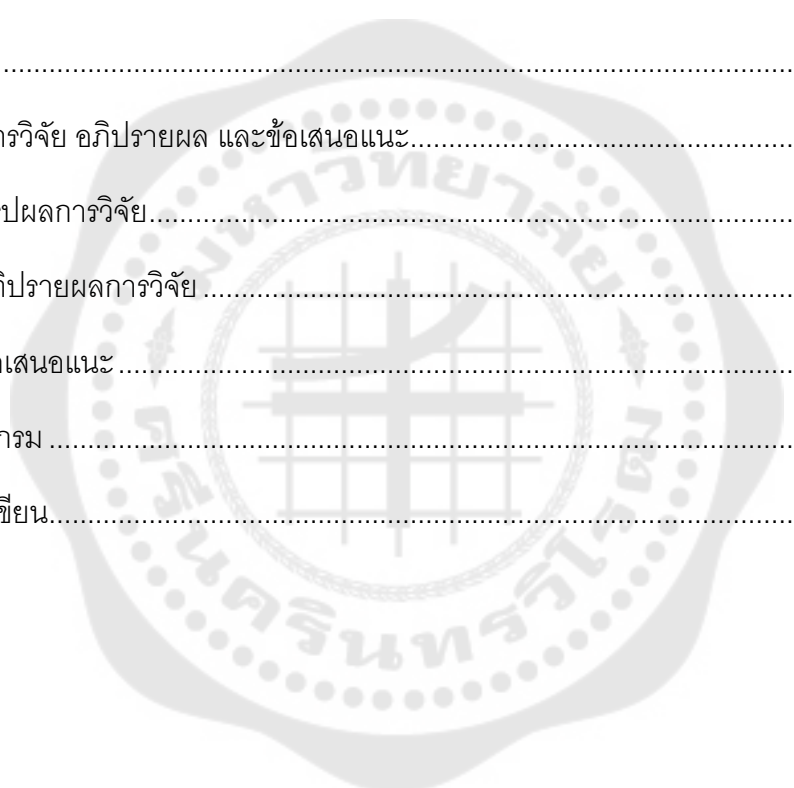
ญาณเดช ภาควัฒรูปร่าง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มาของปัญหาที่ทำการวิจัย	1
1.2 จุดมุ่งหมายของการวิจัย	3
1.3 ขอบเขตการวิจัย	3
1.4 วิธีการดำเนินการวิจัย.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย	5
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	6
2.1 ทฤษฎีที่เกี่ยวข้อง	6
2.1.1 Rate of Penetration (ROP)	6
2.1.2 Random Forest (RF)	13
2.1.3 K-Nearest Neighbor (KNN)	14
2.1.4 Gradient Boosting (GB).....	18
2.1.5 Extreme Gradient Boosting (XGB)	19
2.1.6 Adaptive Boosting (AdaBoost).....	19
2.1.7 Light Gradient Boosting Machine (LGBM).....	21

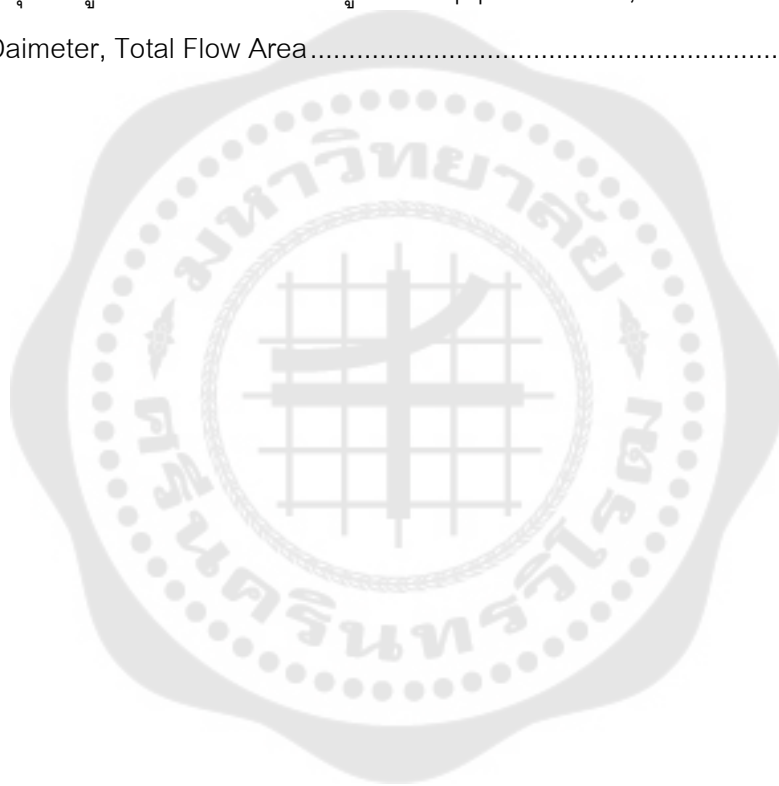
2.1.8 การแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง (Continuous Learning or Forward-Chaining Split) และ แบบหน้าต่างบานเลื่อน (Sliding Window Split)	21
2.1.9 วิธีการประเมินผลแบบจำลอง (Model Evaluation)	25
2.2 งานวิจัยที่เกี่ยวข้อง	28
บทที่ 3 การดำเนินงานวิจัย	41
3.1 กระบวนการทำงานของแบบจำลอง (Model Overview Workflow)	42
3.2. การเตรียมข้อมูล (Data Preparation)	42
3.2.1 ที่มาของแหล่งข้อมูล (Data Source)	42
3.2.2 การได้มาของข้อมูล (Data Acquisition)	46
3.2.3 การทำความสะอาดและเตรียมข้อมูล (Data Cleaning and Preprocessing)	51
3.2.4 การเลือกค่าพารามิเตอร์ที่เกี่ยวข้อง (Selection Parameters)	55
3.3 การสำรวจข้อมูลคุณลักษณะเบื้องต้น (Exploratory Data Analysis)	56
3.3.1 ข้อมูลที่ผิดปกติ (Outliers)	56
3.3.2 การหาค่าความสัมพันธ์ระหว่างตัวแปร (Parameters Correlation)	60
3.3.3 การอธิบายค่าตัวแปร (Explain Feature)	60
3.4 การตั้งค่าแบบจำลองและการฝึกฝน (Model setup and Training)	63
3.4.1 การแบ่งชุดข้อมูลการฝึกฝนและการทดสอบ (Train-Test Split)	63
3.4.2 การตั้งค่าแบบจำลอง (Model setup)	65
3.5 วิธีการประเมินผลแบบจำลอง (Model Evaluation)	66
บทที่ 4	67
ผลการดำเนินงานวิจัย	67
4.1 ผลลัพธ์การเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) ของการแบบจำลองต่างๆ	67

4.2 ผลลัพธ์การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ของการแบบจำลอง ต่างๆ.....	69
4.3 ผลลัพธ์เปรียบเทียบประสิทธิภาพการเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยค่า MAE ของแต่ละ Increments.....	71
4.4 ผลลัพธ์แบบจำลองในการทำนายทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะ จริง.....	72
บทที่ 5.....	74
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	74
5.1 สรุปผลการวิจัย.....	74
5.2 อภิปรายผลการวิจัย.....	75
5.3 ข้อเสนอแนะ.....	79
บรรณานุกรม.....	80
ประวัติผู้เขียน.....	84



สารบัญตาราง

	หน้า
ตาราง 1 อธิบายนิยามคำศัพท์เฉพาะ	12
ตาราง 2 สรุปข้อมูลรวมทางสถิติของข้อมูล Rate of Penetration, Measure Depth, Weight on Bit, Rotary Speed, Surface Torque	58
ตาราง 3 สรุปข้อมูลรวมทางสถิติของข้อมูล Standpipe Pressure, Mud Flow In, Mud Weight out, Bit Diameter, Total Flow Area	59



สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1	1
ภาพประกอบ 2 ภาพแทนเจาะบนบก	2
ภาพประกอบ 3 แสดงขั้นตอนวิธีการดำเนินงานวิจัย (Overview Flowchart).....	4
ภาพประกอบ 4 แสดงองค์ประกอบของแทนเจาะ	7
ภาพประกอบ 5 แสดงปัจจัยที่ส่งผลต่ออัตราการเจาะ	11
ภาพประกอบ 6 หลักการทำงานของ Decision Tree	13
ภาพประกอบ 7 หลักการทำงานของ Random Forest	14
ภาพประกอบ 8 แสดงแนวคิดของแบบจำลอง KNN.....	15
ภาพประกอบ 9 แสดงแนวคิดการหาความคล้ายหรือระยะห่างระหว่างจุดข้อมูลของ KNN	15
ภาพประกอบ 10 แสดงการหาระยะทางโดย Euclidean Distance	16
ภาพประกอบ 11 แสดงการหาระยะทางโดย Manhattan Distance	17
ภาพประกอบ 12 แสดงโครงสร้าง Gradient Boosting Decision Tree	18
ภาพประกอบ 13 แสดงแนวโน้มความผิดพลาดกับแบบจำลองที่เพิ่มขึ้น	19
ภาพประกอบ 14 แสดงโครงสร้าง Adaptive Boosting (AdaBoost).....	20
ภาพประกอบ 15 แสดงโครงสร้าง Adaptive Boosting (AdaBoost).....	20
ภาพประกอบ 16 แสดงโครงสร้างแบบจำลอง LGBM ด้วยเทคนิค Leaf-Wise Tree Growth	21
ภาพประกอบ 17 แสดง K fold cross validation.....	22
ภาพประกอบ 18 แสดง K fold ในแต่ละรอบการวนซ้ำ (Iteration)	22
ภาพประกอบ 19 แสดง Continuous Learning หรือ Forward-Chaining Split และ Sliding Window Split.....	23
ภาพประกอบ 20 แสดง Continuous Learning หรือ Forward-Chaining Split	23

ภาพประกอบ 21 แสดง Continuous Learning หรือ Forward-Chaining Split	24
ภาพประกอบ 22 แสดง Sliding Window Split	24
ภาพประกอบ 23 แสดง Mean Absolute Error (MAE)	26
ภาพประกอบ 24 แสดง Linear Regression	27
ภาพประกอบ 25 ปัจจัยที่มีผลต่อการทำนายอัตราการเจาะ	28
ภาพประกอบ 26 แสดงตารางค่าทางสถิติของพารามิเตอร์ต่างๆ	29
ภาพประกอบ 27 แสดงตารางผลของแบบจำลอง KNN และ MLP	30
ภาพประกอบ 28 แสดงกราฟผลของแบบจำลอง KNN และ MLP	30
ภาพประกอบ 29 แสดงการเปรียบเทียบเทียบของการทำนายอัตราการเจาะกับอัตราการเจาะจริงในอดีตในช่วงของระดับความลึกต่างๆในแบบจำลอง KNN และ MLP	31
ภาพประกอบ 30 แสดงตารางหมวดหินใน Williston basin, North Dakota.....	32
ภาพประกอบ 31 แสดงความสำคัญของพารามิเตอร์	33
ภาพประกอบ 32 แสดงตารางผล R2 และ RMSE ของการทำนายอัตราการเจาะในแบบจำลองต่างๆ.....	34
ภาพประกอบ 33 แสดงการเปรียบเทียบข้อมูลการทดสอบ (Test data), ข้อมูลการฝึกสอน (Training data) และข้อมูลการทำนายอัตราการเจาะ (Rate Of Penetration , ROP) ในแบบจำลอง Random forest และ Bingham	35
ภาพประกอบ 34 แสดงประสิทธิภาพของแบบจำลองด้วย Default model และ Genetic Algorithms.....	37
ภาพประกอบ 35 แสดงการเปรียบเทียบปัจจัยการขุดเจาะใน 4 แบบจำลองเชิงวิเคราะห์ ROP ..	38
ภาพประกอบ 36 แสดงผลการแบ่งข้อมูลเพื่อการฝึกฝนและการทดสอบ	40
ภาพประกอบ 37 แสดงกระบวนการทำงานของแบบจำลอง (Model Overview Workflow)	42
ภาพประกอบ 38 แสดงหน้า website บริษัท Equinor ที่สามารถดาวน์โหลดข้อมูลได้.....	43

ภาพประกอบ 39 แสดงพื้นที่บริเวณไหล่ทวีปของนอร์เวย์ (Norwegian Continental shelf : NCS)	44
.....	
ภาพประกอบ 40 แสดงจุดพื้นที่ของ Volve filed ในกลางทะเลเหนือ	45
ภาพประกอบ 41 แสดงแหล่งพื้นที่ของ Volve filed	45
ภาพประกอบ 42 แสดงการเข้าถึงการดาวน์โหลดข้อมูล	46
ภาพประกอบ 43 แสดงหน้าการ Login B2C	46
ภาพประกอบ 44 แสดงชื่อชุดข้อมูล Volve Data Village	47
ภาพประกอบ 45 แสดง Azure signature URL	47
ภาพประกอบ 46 แสดง Azure signature URL	48
ภาพประกอบ 47 แสดง Azure signature	48
ภาพประกอบ 48 แสดง Azure signature URL	49
ภาพประกอบ 49 แสดง Azure signature URL	49
ภาพประกอบ 50 แสดงชุดข้อมูลทั้งหมดภายใน Volve Data Village	50
ภาพประกอบ 51 แสดงทิศทางการเจาะของหลุมเจาะเอียงไปจากแนวตรง (Deviation Well)	51
ภาพประกอบ 52 แสดงการคัดเลือกคอลัมน์ในเบื้องต้น	52
ภาพประกอบ 53 แสดงการตัดช่วงความลึกที่ไม่ได้มีการเก็บข้อมูลไว้ออกจากข้อมูลทั้งหมด	52
ภาพประกอบ 54 แสดงการแยกข้อมูลตัวหนังสือและตัวเลขออกจากกันและแทนที่ด้วยค่าตัวเลขเพียงอย่างเดียว	53
ภาพประกอบ 55 แสดงการเติมค่าว่าง (NaN) ทั้งหมดในข้อมูล	53
ภาพประกอบ 56 แสดงค่าติดลบที่อยู่ในข้อมูล	54
ภาพประกอบ 57 แสดงค่าติดลบที่อยู่ในข้อมูล	54
ภาพประกอบ 58 แสดงการเพิ่มคอลัมน์ Total Flow Area	55
ภาพประกอบ 59 แสดงรายละเอียดข้อมูลคอลัมน์ที่สำคัญและเกี่ยวข้อง	55
ภาพประกอบ 60 แสดงรายละเอียดข้อมูลเบื้องต้นทั้งหมด	56

ภาพประกอบ 61 แสดงค่าความผิดพลาดของค่าอัตราการเจาะ.....	57
ภาพประกอบ 62 แสดงค่าความผิดพลาดของค่า Rotary Speed.....	57
ภาพประกอบ 63 แผนภูมิความร้อนแสดงการหาค่าความสัมพันธ์ระหว่างตัวแปร	60
ภาพประกอบ 64 การอธิบายเปรียบเทียบปัจจัยกับช่วงค่าความลึกต่างๆที่วัดได้ (Measure Depth, MD) (ก) ค่าอัตราการเจาะ (Rate of Penetration, ROP) (ข) ค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) (ค) ค่าอัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM) (ง) ค่าน้ำหนักน้ำโคลนที่ออกมา (Mud Weight out, MWout) (จ) ค่าความดันที่ก้านเจาะ (Standpipe Pressure, SPP) (ฉ) การไหลเข้าของน้ำโคลน (Mud Flow In, Q) (ญ) พื้นที่หน้าตัดรวมที่ทำให้ของไหลผ่านเข้าไปในหัวเจาะ (Total Flow Area, TFA) (ฐ) แรงบิดตัวด้านบน (Surface Torque, T) (ถ) ขนาดของหัวเจาะ (Bit Diameter, Dbit)	63
ภาพประกอบ 65 การแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or forward-chaining split)	65
ภาพประกอบ 66 การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split).....	65
ภาพประกอบ 67 แสดงค่า MAE ของการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) ใน แบบจำลองต่างๆ	68
ภาพประกอบ 68 แสดงกราฟแท่งการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) แต่ละ increments ในการเปรียบเทียบแบบจำลองต่างๆกับค่า MAE.....	68
ภาพประกอบ 69 แสดงค่า MAE ของการเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ใน แบบจำลองต่างๆ.....	70
ภาพประกอบ 70 แสดงกราฟแท่งการเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) แต่ละ increments ในการเปรียบเทียบแบบจำลองต่างๆกับค่า MAE.....	70
ภาพประกอบ 71 แสดงการเปรียบเทียบระหว่างประสิทธิภาพการเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยค่า MAE ของแต่ละ Increments	72
ภาพประกอบ 72 แสดงการทำนายทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะจริง ในแบบจำลอง XGB	72

ภาพประกอบ 73 ค่าความคลาดเคลื่อนที่ยอมรับได้ไม่เกิน 3 เท่าของส่วนเบี่ยงเบนมาตรฐาน....77



บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

ในช่วง 2-3 ทศวรรษที่ผ่านมาพลังงานเป็นสิ่งหนึ่งที่สำคัญในกิจกรรมต่างๆของมนุษย์ โดยเฉพาะพลังงานฟอสซิลที่เป็นแหล่งพลังงานด้านปิโตรเลียมคือน้ำมันและก๊าซ เนื่องจากอัตราการใช้ที่เพิ่มสูงขึ้นอย่างต่อเนื่องในช่วงหลายปีที่ผ่านมา ทำให้การขุดเจาะหาน้ำมันและก๊าซเพื่อตอบสนองได้อย่างเพียงพอต่อผู้คนที่ทั่วโลกและความต้องการของตลาด ทำให้การลงทุนในการขุดเจาะของบริษัทน้ำมันต่างๆทั้งบนบกและในทะเล ดังภาพประกอบที่ 1 และ 2 ต้องคำนึงถึงค่าใช้จ่ายในการขุดเจาะเพื่อแสวงหาแหล่งน้ำมันและก๊าซนั้น โดยมีการลงทุนที่มีมูลค่ามากต่อ 1 หลุมเจาะ ไม่ว่าจะเป็นการลงทุนด้านทรัพยากรบุคคล อุปกรณ์และเครื่องมือต่างๆที่ใช้ในการขุดเจาะปิโตรเลียม ดังนั้นการลดค่าใช้จ่ายในการขุดเจาะปิโตรเลียมจึงเป็นส่วนสำคัญในการขุดเจาะเพื่อเป็นการลดต้นทุนค่าใช้จ่ายในการขุดเจาะ โดยการขุดเจาะปิโตรเลียมนั้นจะมีตัวแปรในการขุดเจาะอยู่หลายตัวแปร ซึ่งตัวแปรหนึ่งที่เป็นปัจจัยที่มีส่วนสำคัญในการขุดเจาะและช่วยการลดต้นทุน คือ ค่าอัตราการเจาะ (Rate of Penetration , ROP) ซึ่งถ้าสามารถหาค่า ROP ของหลุมเจาะได้อย่างถูกต้อง จะช่วยในการวางแผนระยะเวลาและลดค่าใช้จ่ายในการขุดเจาะปิโตรเลียมได้ หรือแม้กระทั่งเป็นตัวช่วยเพื่อให้วิศวกรขุดเจาะมีตัวช่วยเพื่อประกอบการตัดสินใจในหน้าที่มีการขุดเจาะอยู่นั้น



ภาพประกอบ 1

ที่มา : <https://www.offshore-energy.biz/vantage-drilling-secures-more-jack-up-rig-work/>



ภาพประกอบ 2 ภาพแท่นเจาะบนบก

ที่มา : <https://www.offshore-technology.com/projects/tilapia-oil-field/>

ในปัจจุบันนี้องค์ความรู้ทางวิทยาการทางข้อมูล (Data Science) เป็นที่แพร่หลายไปในหลากหลายอุตสาหกรรมรวมถึงอุตสาหกรรมการขุดเจาะปิโตรเลียม การได้เห็นความสำคัญของการนำทั้ง 2 ศาสตร์มาประยุกต์เข้าด้วยกัน จะช่วยให้การทำงานมีประสิทธิภาพมากขึ้น รวมทั้งเป็นแนวทางทำให้มีกระบวนการเพิ่มประสิทธิภาพของการทำงานด้านการขุดเจาะปิโตรเลียมโดยการนำเอาองค์ความรู้และเทคโนโลยีด้านวิทยาการทางข้อมูลเข้าผสมผสานเพื่อให้เกิดคุณประโยชน์มากขึ้น แต่การทำความเข้าใจในตัวแปรต่างๆและนำค่าตัวแปรในการขุดเจาะมาประยุกต์ใช้นั้น ยังต้องอาศัยความรู้ความเข้าใจในข้อมูลเฉพาะด้านซึ่งเป็นส่วนสำคัญในการวิเคราะห์ข้อมูลได้อย่างถูกต้อง จึงจะสามารถนำวิธีทางวิทยาการทางข้อมูลมารวมกับข้อมูลหลุมเจาะ เพื่อประยุกต์ในการนำไปใช้งานให้เกิดประโยชน์ได้อย่างสูงสุด และก่อให้เกิดประโยชน์ต่อวงการอุตสาหกรรมการขุดเจาะปิโตรเลียมและบริษัทน้ำมันต่างๆ ทั้งยังช่วยบุคลากรในหน่วยงานประกอบการตัดสินใจ รวมถึงช่วยควบคุมค่าใช้จ่ายเพื่อประกอบการวางแผนระยะการขุดเจาะได้อย่างมีประสิทธิภาพ

1.2 จุดมุ่งหมายของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งจุดมุ่งหมายไว้ดังนี้

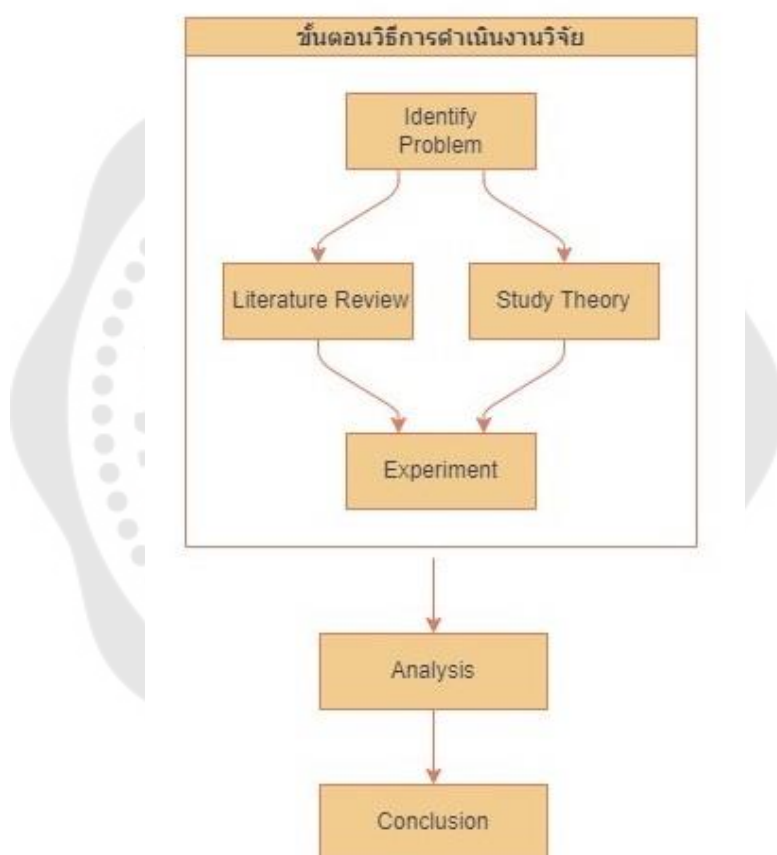
1. เพื่อศึกษาเปรียบเทียบเทคนิคลำดับของข้อมูล (Sequential Data) โดยการแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split)
2. เพื่อศึกษาเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และเปรียบเทียบการทำงานเพื่อวัดประสิทธิภาพในแต่ละอัลกอริทึม (Algorithms) โดยแบบจำลอง (Model) ที่ใช้คือ Random Forest (RF), K-Nearest Neighbor (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost) และ Light Gradient Boosting Machine (LGBM)
3. เพื่อศึกษาเปรียบเทียบการทำนายค่าอัตราการเจาะกับค่าอัตราการเจาะจริงของหลุมเจาะปิโตรเลียมที่เจาะแล้วในอดีตในแบบจำลองที่มีประสิทธิภาพ

1.3 ขอบเขตการวิจัย

งานวิจัยนี้ใช้ข้อมูลหลุมเจาะปิโตรเลียมจากบริษัทน้ำมัน Equinor ของประเทศนอร์เวย์ โดยข้อมูลชื่อ Volve dataset 2018 ซึ่งเป็นข้อมูลการเจาะหลุมสำรวจ (Exploration Well) แบบ Real-time Drilling ที่เผยแพร่เพื่อการศึกษาและวิจัยเท่านั้น ข้อมูลที่ใช้ในการวิเคราะห์มีทั้งหมด 2,701 ข้อมูล ประกอบด้วย 10 ตัวแปร โดยเป้าหมายที่ทำการศึกษา คือ ทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะจริงในอดีต โดยแบบจำลองที่ใช้ ได้แก่ Random Forest (RF), K-Nearest Neighbor (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost) และ Light Gradient Boosting Machine (LGBM) เพื่อนำมาเปรียบเทียบประเมินประสิทธิภาพของแบบจำลองแบบทำนายได้ (Predictive Model) ของการทำงานในแต่ละอัลกอริทึมด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) โดยจะทำการเลือกคุณลักษณะตัวแปรที่สำคัญที่ส่งผลต่อค่าอัตราการเจาะที่ทำให้การทำนายผลแม่นยำมากขึ้น เนื่องจากงานวิจัยนี้เป็นการศึกษาปัญหาด้านการวิเคราะห์ถดถอยปัญหาเชิงเส้น (Linear Regression) จึงใช้วัดค่าประสิทธิภาพด้วย ค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2)

1.4 วิธีการดำเนินการวิจัย

การศึกษาวิจัยนี้อยู่ในแนวทางของการใช้ศาสตร์ของวิทยาการข้อมูล (Data Science) ในการใช้ Machine Learning โดยการเปรียบเทียบแบบจำลอง Random Forest (RF), K-Nearest Neighbor (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost) และ Light Gradient Boosting Machine (LGBM) เพื่อหาค่าการทำนาย อัตราการเจาะของหลุมเจาะปิโตรเลียม โดยมีวิธีขั้นตอนการดำเนินงานวิจัย ดังภาพประกอบ 3



ภาพประกอบ 3 แสดงขั้นตอนวิธีการดำเนินงานวิจัย (Overview Flowchart)

1.5 ประโยชน์ที่คาดว่าจะได้รับการวิจัย

- 1.สามารถนำแบบจำลองที่ได้ไปช่วยทำนายค่าอัตราการเจาะ เพื่อใช้ประกอบในการวางแผนการดำเนินงานเบื้องต้นได้
- 2.สามารถนำค่าอัตราการเจาะที่ทำนายได้ ไปใช้ในการประเมินเวลาการเจาะเพื่อควบคุมค่าใช้จ่ายในการดำเนินงานได้
- 3.สามารถนำค่าอัตราการเจาะที่ทำนายได้ ไปเป็นแนวทางให้วิศวกรขุดเจาะ (Drilling Engineer) ประกอบการตัดสินใจในหน้างานขณะขุดเจาะได้



บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องเกี่ยวกับการเปรียบเทียบแบบจำลองในการทำนายค่าอัตราการเจาะของหลุมปิโตรเลียม และได้เสนอตามหัวข้อดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 Rate of Penetration (ROP)

2.1.2 Random Forest (RF)

2.1.3 K-Nearest Neighbor (KNN)

2.1.4 Gradient Boosting (GB)

2.1.5 Extreme Gradient Boosting (XGB)

2.1.6 Adaptive Boosting (AdaBoost)

2.1.7 Light Gradient Boosting Machine (LGBM)

2.1.8 การแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ แบบหน้าต่างบานเลื่อน (Sliding Window Split)

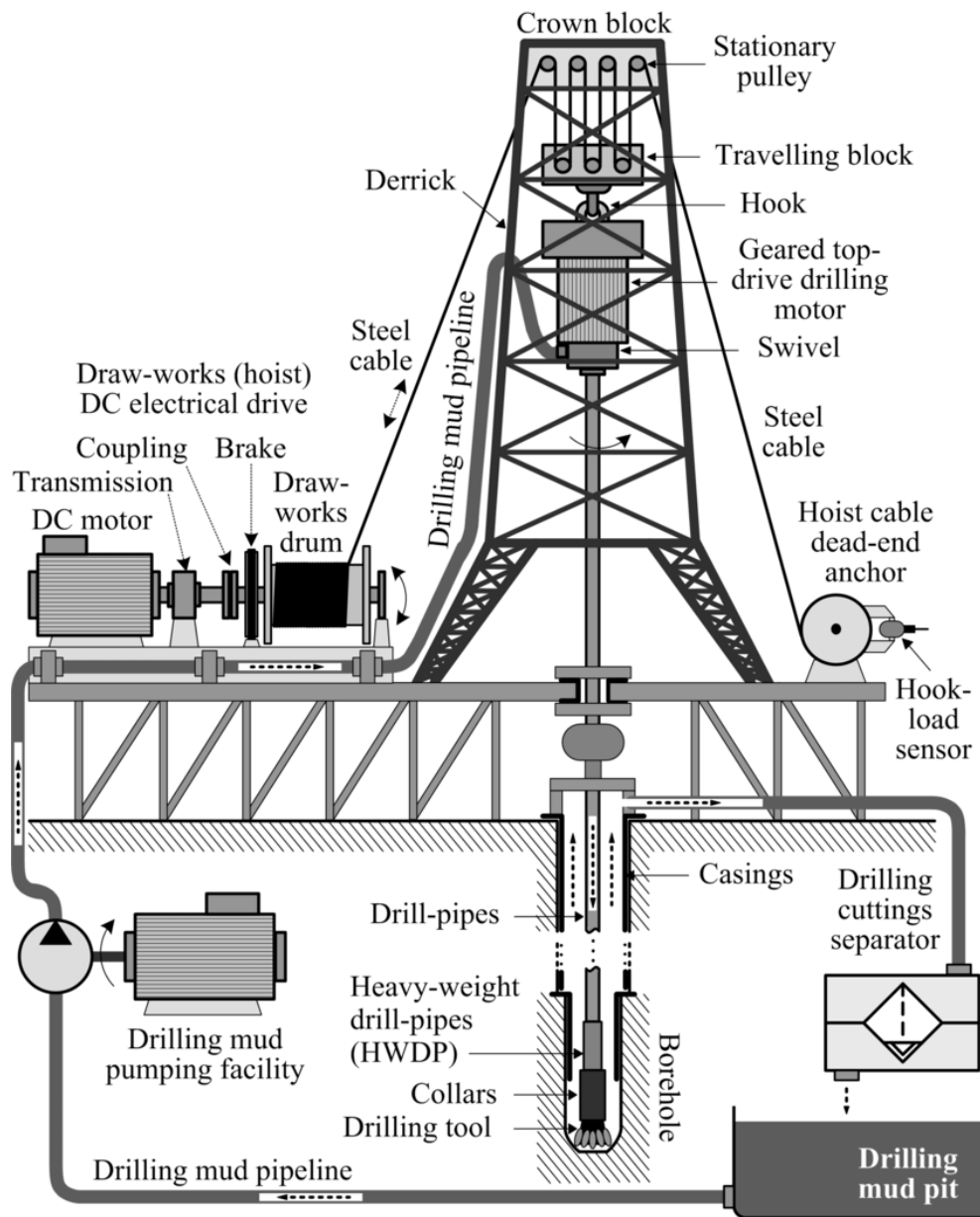
2.1.9 วิธีการประเมินผลแบบจำลอง (Model Evaluation)

2.2 งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 Rate of Penetration (ROP)

การขุดเจาะปิโตรเลียมมีทั้งบนบก (Onshore) และ ในทะเล (Offshore) ซึ่งแท่นเจาะมีองค์ประกอบที่สำคัญเหมือนกัน ดังภาพประกอบ 4 ในการขุดเจาะมีปัจจัยหลายอย่างที่เกี่ยวข้อง โดยปัจจัยหนึ่งที่สำคัญ คือ อัตราการเจาะ (Rate Of Penetration, ROP)



ภาพประกอบ 4 แสดงองค์ประกอบของแท่นเจาะ

ที่มา : Sprljan et al. (2020). Automation Systems Design and Laboratory Prototyping Aimed at Mature Petroleum Drilling Rig Retrofitting

การคำนวณสูตรของแบบจำลองแบบเดิม (Traditional Model) จะอาศัยแบบจำลองทางฟิสิกส์โดยใช้สูตรและฟังก์ชันทางคณิตศาสตร์ แบบจำลอง Speer (Speer, 1959) ได้มีการหาความสัมพันธ์ระหว่าง อัตราการเจาะ (Rate Of Penetration , ROP), น้ำหนักในการกดลงบนหัวเจาะ (Weight On Bit, WOB) และ ความเร็วรอบในการหมุน (Rotary Speed : Revolution per minute, RPM) ซึ่งจะรู้จักกันในความสัมพันธ์ของ R-W-N เนื่องจาก ROP เกิดจาก WOB และ RPM แบบจำลอง Maurer (Maurer, 1962) ได้พัฒนาจากทฤษฎีการคำนวณสำหรับ rolling cutting bits (Maurer, 1962, apud (Bourgoyne et al., 1986), p. 226) โดยให้การทำความสะดวกที่ก้นหลุมแบบสมบูรณ์ (Perfect bottomhole cleansing) และความไม่สมบูรณ์ของฟันของหัวเจาะ (incomplete bit tooth penetration) แสดงดังสมการที่ (1)

$$ROP = \frac{K}{S^2} \left[\frac{W}{d_b} - \frac{W_0}{d_b} \right]^2 N$$

(1)

โดยที่

K คือ constant of proportionality

S คือ compressive rock strength

W คือ bit weight

 W_0 คือ threshold bit weight d_b คือ drill bit diameter

N คือ rotary speed

อย่างไรก็ตามแบบจำลอง Maurer ไม่สามารถประยุกต์ใช้ได้ทั่วไป เนื่องจากเส้นโค้ง (curve) ที่ตอบสนองของ R-W-N ให้ความแตกต่างกัน ยังมีแบบจำลองที่เสนอโดย Bingham (Bingham, 1964) ได้ให้ความสัมพันธ์ระหว่าง ROP, WOB, RPM และ Bit Diameter โดยการเพิ่มเลขชี้กำลังเข้าไปเทอมของ WOB (Bingham, 1965, apud (Bourgoyne et al., 1986), p. 227) แสดงดังสมการที่ (2) ซึ่งยังรวมถึงการเกี่ยวข้อของของ Formation ของชั้นหิน กับ WOB ที่เป็นเลขชี้กำลังของแบบจำลอง ROP ที่เสนอโดย Murray และ Cunningham (Murray & Cunningham, 1955) เป็นครั้งแรก ซึ่งเป็นการคำนวณในหินแต่ละกลุ่ม (Rock Formation)

$$ROP = a \left(\frac{WOB}{d_b} \right)^b RPM \quad (2)$$

โดยที่

a และ b คือ dimensionless constants for each formation

WOB คือ bit weight

d_b คือ drill bit diameter

RPM คือ rotational speed

Bourgoyne และ Young ((Bourgoyne & Young, 1974) ได้พัฒนาแบบจำลอง ROP โดยอธิบายถึงความแตกต่างกันของ 8 Parameters แสดงดังสมการที่ (3)

$$ROP = \text{Exp} \left(a_1 + \sum_{j=2}^8 a_j x_j \right) \quad (3)$$

โดยที่

a_1 คือ the formation strength parameter

a_2 คือ the normal compaction trend exponent

a_3 คือ the undercompaction exponent

a_4 คือ the pressure differential exponent

a_5 คือ the bit weight exponent

a_6 คือ the rotational speed exponent

a_7 คือ the tooth wear exponent

a_8 คือ the hydraulic exponent

แบบจำลองของ Bourgoyne และ Young (Bourgoyne & Young, 1974) เป็นแบบจำลอง ROP พื้นฐานที่ได้รับความนิยมในการพัฒนาสำหรับ Roller cone bit รวมถึงยังเป็นที่แพร่หลายในการประยุกต์ใช้กับการเจาะ PDC Bit หลังจากนั้น Bourgoyne และคณะ

(Bourgoyne et al., 1986) ได้พัฒนาแบบจำลองที่มีการปรับตัวแปรและความลึกของหลุมให้เข้ากับการเจาะแบบ Real-time

แบบจำลอง PDC bit ได้มาจาก Hareland และ Rampersad (Hareland & Rampersad, 1994) และ Motahhari และคณะ (Motahhari et al., 2010) ซึ่งจะเกี่ยวข้องกับความสัมพันธ์ของ Cutters และ ความเร็วในการเจาะของแต่ละกลุ่มหิน โดย Hareland และ Rampersad (1994, apud (Soares et al., 2016), p. 1229) ได้เสนอแบบจำลองของ Drag bit แบบทั่วไป แสดงดังสมการที่ (4)

$$ROP = W_f \frac{a}{(RPM)^b WOB^c} \frac{14.14 N_c RPM}{d_b} \cos \alpha \sin \theta \left[\left(\frac{d_c}{2} \right)^2 \cos^{-1} \left(1 - \frac{4 W_{mech}}{\pi \cos \theta d_c^2 \sigma_c} \right) - \left(\frac{2 W_{mech}}{\pi \cos \theta \sigma_c} - \frac{4 W_{mech}^2}{(\pi \cos \theta d_c \sigma_c)^2} \right)^{0.5} \left(\frac{d_c}{2} - \frac{W_{mech}}{\pi \cos \theta d_c \sigma_c} \right) \right] \quad (4)$$

โดยที่

a, b, และ c คือ cutter geometry correction factors

Wf คือ a bit wear function

Nc คือ the number of PDC cutters

α คือ the cutter side rake angle,

θ คือ the cutter back rake angle

dc คือ the cutter diameter (in)

Wmech คือ the mechanical loading per cutter (lbf)

σ_c คือ the uniaxial (or unconfined) compressive rock strength

(UCS) (psi)

ในความเป็นจริงค่าความสัมพันธ์ของพารามิเตอร์ในการเจาะมีความซับซ้อนและเข้าใจได้ยาก ดังนั้นจึงมีการอธิบายถึงความสัมพันธ์ของพารามิเตอร์ที่ส่งผลต่อการเจาะ โดย Motahhari และคณะ (Motahhari et al., 2010) ได้เสนอแบบจำลอง ROP ของ PDC bit แสดงดังสมการที่ (5)

$$ROP = W_f \left(\frac{G \text{ RPM}^\gamma \text{ WOB}^\alpha}{d_b S} \right) \quad (5)$$

โดยที่

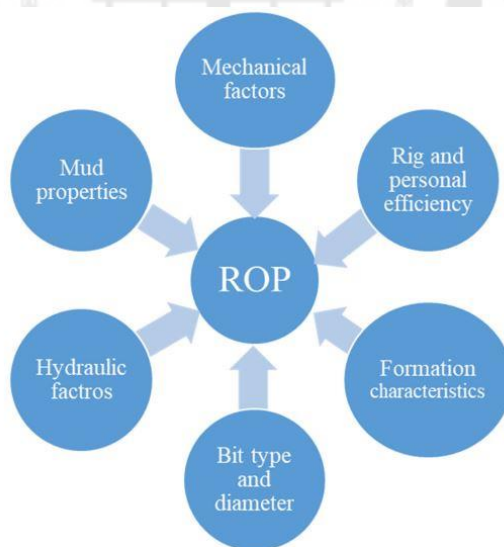
W_f คือ a bit wear function

G คือ a coefficient related to bit-rock interactions และ bit geometry

α และ γ คือ ROP model exponents

S คือ the confined compressive rock strength (CCS) [psi]

จากการพัฒนาแบบจำลองแบบเดิมในการหาอัตราการเจาะจากที่กล่าวมา มีปัจจัยที่กำหนดในแต่ละแบบจำลองที่แตกต่างกันออกไป โดยปัจจัยที่มีผลสามารถแบ่งเป็นกลุ่มใหญ่ๆจะแบ่งได้แก่ 1.Mechanical factors 2.Mud properties 3.Hydraulic factors 4.Bit type and diameter 5.Formation characteristics 6.Rig and personal efficiency ดังภาพประกอบ 5



ภาพประกอบ 5 แสดงปัจจัยที่ส่งผลต่ออัตราการเจาะ

ที่มา : El-Sayed et al.,(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

ตาราง 1 อธิบายนิยามคำศัพท์เฉพาะ

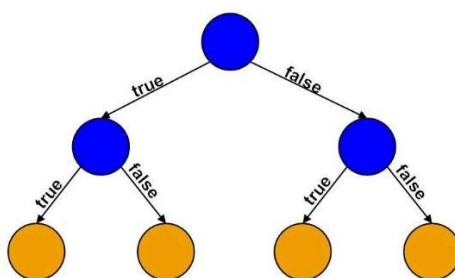
Feature	Definition	Unit
Rate of Penetration (ROP)	อัตราเร็วในการเจาะ	Meter / hour (m/hr)
Measure Depth (MD)	วัดได้ลึกน้อยไปตามแนวหลุม	Meter (m)
Weight on Bit (WOB)	น้ำหนักที่กดลงบนหัวเจาะ	Kilogram force (kggf)
Rotary Speed (RPM)	ความเร็วรอบของก้านเจาะ	Revolution per minute (rpm)
Surface Torque (T)	แรงบิดตัวด้านบน	Kilonewtons. Meter (kN.m)
Standpipe Pressure (SPP)	ความดันที่ก้านเจาะ	Kilopascal (kPa)
Mud Flow In (Q)	การไหลเข้าของน้ำโคลน	Liter / minute (L/min)
Mud Weight Out (MW _{out})	น้ำหนักของน้ำโคลนที่ออกมา	Grams / cubic centimeter (g/cm ³)
Bit Diameter (D _{BIT})	ขนาดของหัวเจาะ	Inch
Total Flow Area (TFA)	พื้นที่หน้ารวมที่ของไหลผ่านเข้าไปในหัวเจาะ	Inch ²

การวิจัยนี้จะศึกษาในรูปแบบของการถดถอย (Regression) โดยแบบจำลองที่ศึกษาจะประกอบไปด้วย

- 1, Bagging (Tree-based) ensemble model, or Bootstrap aggregation ได้แก่ Random Forest
2. Instance-based model ได้แก่ K-Nearest Neighbor
3. Boosting ensemble model ได้แก่ Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LGBM)

2.1.2 Random Forest (RF)

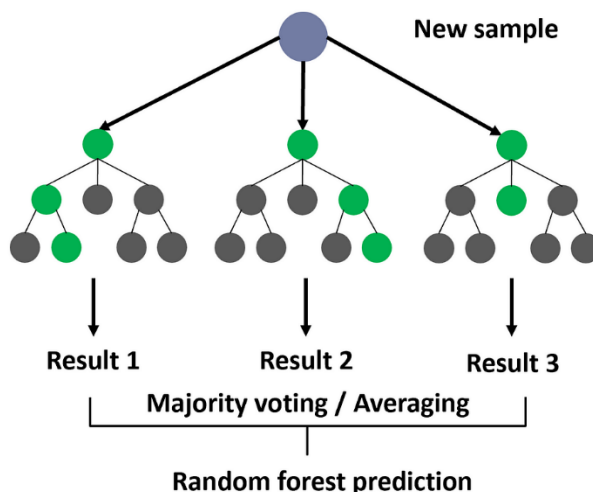
ทฤษฎีการสุ่มแบบป่าไม้ (Random Forest) เป็นแบบจำลองแบบ Supervise Learning ที่อยู่ในกลุ่มแบบจำลอง Ensemble Learning โดยรูปแบบจำลองถูกพัฒนามาจาก Decision tree ซึ่งเป็นการเพิ่มจำนวนต้นไม้ให้เยอะขึ้นแบบมีลำดับชั้น โดยตัดสินใจแบบ if-else ดังภาพประกอบ 6 ถูกนำเสนอครั้งแรกในปี 1995 โดย Tim Kam (Tin Kam Ho, 1995) ในบทความที่ชื่อว่า Random Decision Forests และต่อมาถูกพัฒนาโดย Leo Breiman และ Adele Cutler ในบทความที่ชื่อว่า Random Forests ที่เขียนโดย Leo Breiman (Breiman, 2001) และเผยแพร่ในปี 2001



ภาพประกอบ 6 หลักการทำงานของ Decision Tree

ที่มา: https://help.sap.com/saphelp_SNC700_ehp01/helpdata/en/99/02f1afe99c46cda61f1363755101e9/content.htm?no_cache=true

ส่วน Random Forest หลักการ คือ การฝึกฝนชุดข้อมูลบนแบบจำลองที่เหมือนกันหลายๆครั้ง บนชุดข้อมูลเดียวกัน ในแต่ละครั้งของการฝึกฝนชุดข้อมูลจะเลือกส่วนของข้อมูลที่ฝึกฝนไม่เหมือนกัน แล้วนำการตัดสินใจของแบบจำลองมาทำการโหวต class ที่มากที่สุด ดังภาพประกอบ 7

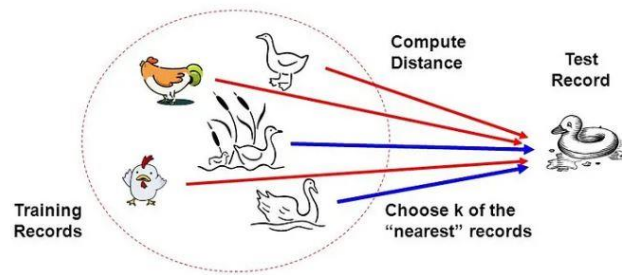


ภาพประกอบ 7 หลักการทำงานของ Random Forest

ที่มา : <https://medium.com/@roiyehe/random-forests-98892261dc49>

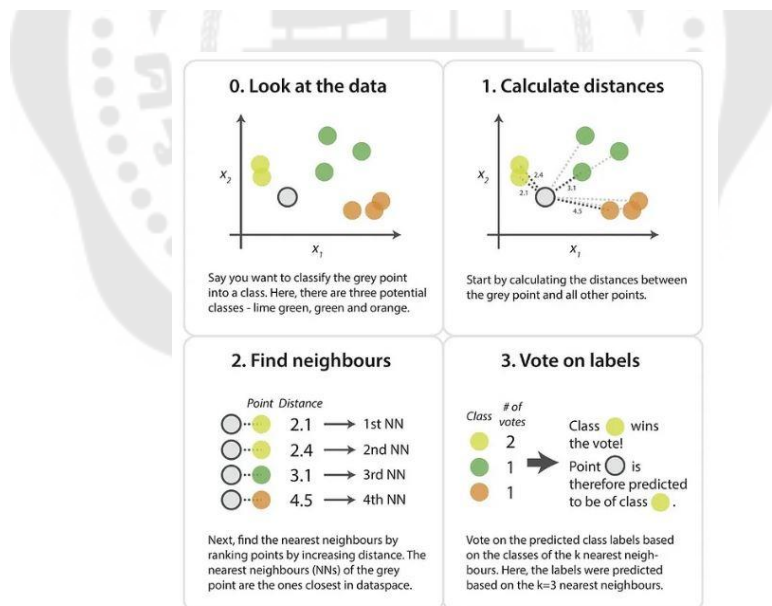
2.1.3 K-Nearest Neighbor (KNN)

Evelyn Fix และ Joseph Hodges (Fix & Hodges, 1951) เป็นริเริ่มผู้ให้แนวคิดเกี่ยวกับ K-Nearest Neighbor หรือแบบจำลอง KNN ในปี 1951 ขณะที่ Thomas Cover (T. Cover & P. Hart, 1967) ได้อธิบายแนวคิดเพิ่มเติมในบทความ Nearest Neighbor Pattern Classification แบบจำลอง KNN ใช้เทคนิค Instance-based เป็นการเรียนรู้ (Supervised Learning) แบบไม่ใช้พารามิเตอร์ (non-parametric) แต่ใช้การค้นหาความคล้ายหรือระยะห่างระหว่างจุดข้อมูล โดยหลักการที่ว่า ข้อมูลที่คล้ายกันจะอยู่ใกล้กัน ดังภาพประกอบ 8 และ 9 โดยชุดข้อมูลที่ใกล้ที่สุดที่ใช้ในการตัดสินใจ จะเรียกว่า ค่า k ซึ่งในการทำ Regression นั้น KNN จะหาค่าเฉลี่ยของข้อมูลใกล้เคียงเพื่อทำนายค่าตัวเลขสำหรับข้อมูลใหม่ (Olafadehan & Ahaotu, 2023)



ภาพประกอบ 8 แสดงแนวคิดของแบบจำลอง KNN

ที่มา: <https://kongruksiam.medium.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-machine-learning-ep-4-%E0%B9%80%E0%B8%9E%E0%B8%B7%E0%B9%88%E0%B8%AD%E0%B8%99%E0%B8%9A%E0%B9%89%E0%B8%B2%E0%B8%99%E0%B9%83%E0%B8%81%E0%B8%A5%E0%B9%89%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%AA%E0%B8%B8%E0%B8%94-k-nearest-neighbors-787665f7c09d>



ภาพประกอบ 9 แสดงแนวคิดการหาความคล้ายหรือระยะห่างระหว่างจุดข้อมูลของ KNN

ที่มา : <https://towardsdatascience.com/how-to-find-the-optimal-value-of-k-in-knn-35d936e554eb>

การเลือกค่า K ควรพิจารณาดังนี้

1. ค่า K ควรเป็นจำนวนเลขคี่ เพื่อลดในเรื่องการมีผลโหวตที่เท่ากัน
2. ถ้าจำนวนข้อมูลในชุดการฝึกสอนมาก ค่า K ควรมีค่าน้อย เพื่อลดความซับซ้อนและเพิ่มประสิทธิภาพในการคำนวณ
3. ถ้าจำนวนข้อมูลในชุดการฝึกสอนน้อย ค่า K ควรมีค่ามาก เพื่อป้องกันการเอียงในการจัดกลุ่มหรือจำแนกประเภท

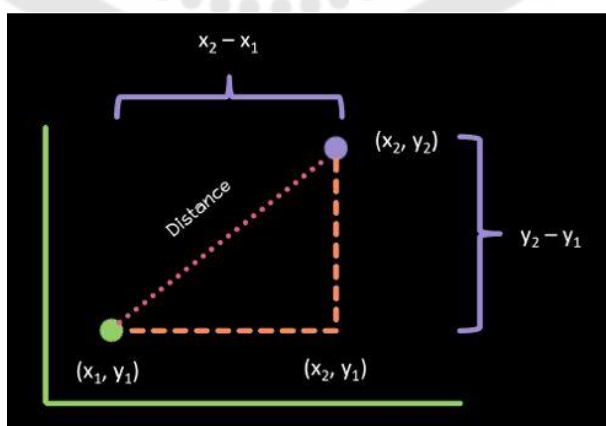
ในการหาระยะทางเป็นการคำนวณระยะห่างระหว่างข้อมูลที่นิยมใช้กันจะมีอยู่ 3 ประเภท คือ

1. Euclidean Distance เป็นวิธีคำนวณระยะห่างโดยใช้สูตรสามเหลี่ยมพีทาโกรัส (Pythagorean theorem) เพื่อหาระยะห่างระหว่างจุดในพื้นที่ของคุณลักษณะ (feature space) ดังภาพประกอบ 10 สูตรการคำนวณ ดังสมการที่ (6)

$$\text{ระยะทาง Euclidean} = \sqrt{(X_2 - X_1)^2 + (y_2 - y_1)^2} \quad (6)$$

โดยที่

(X_1, y_1) และ (X_2, y_2) คือ ค่าของคุณลักษณะของข้อมูลสองจุดที่จะหาระยะห่าง



ภาพประกอบ 10 แสดงการหาระยะทางโดย Euclidean Distance

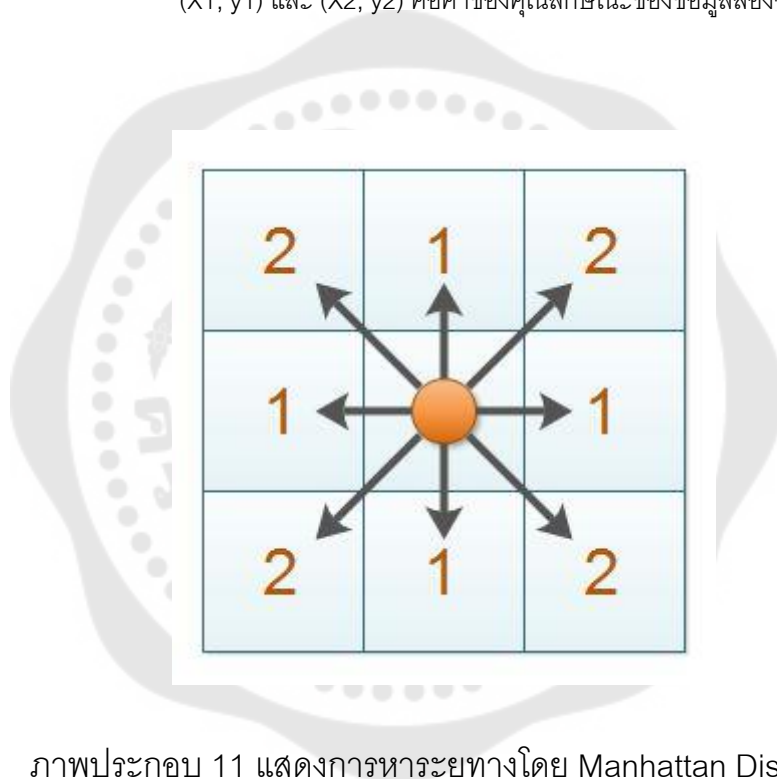
ที่มา : <https://digi.data.go.th/blog/what-is-k-nearest-neighbors/>

2. Manhattan Distance เป็นวิธีคำนวณระยะห่างโดยการบวกค่าของความแตกต่างระหว่างแต่ละคู่ของคุณลักษณะ ดังภาพประกอบ 11 สูตรการคำนวณ ดังสมการที่ (7)

$$\text{ระยะทาง Manhattan} = |X_2 - X_1| + |y_2 - y_1| \quad (7)$$

โดยที่

(X_1, y_1) และ (X_2, y_2) คือค่าของคุณลักษณะของข้อมูลสองจุดที่จะหาระยะห่าง



ภาพประกอบ 11 แสดงการหาระยะทางโดย Manhattan Distance

ที่มา : <https://iq.opengenus.org/manhattan-distance/>

3. Minkowski Distance เป็นวิธีคำนวณระยะห่างที่มีความยืดหยุ่นที่สุด โดยสามารถกำหนดค่า "p" เพื่อปรับระยะห่าง Minkowski สูตรการคำนวณ ดังสมการที่ (8)

$$\text{ระยะทาง Minkowski} = (|X_2 - X_1|^p + |y_2 - y_1|^p)^{1/p} \quad (8)$$

โดยที่

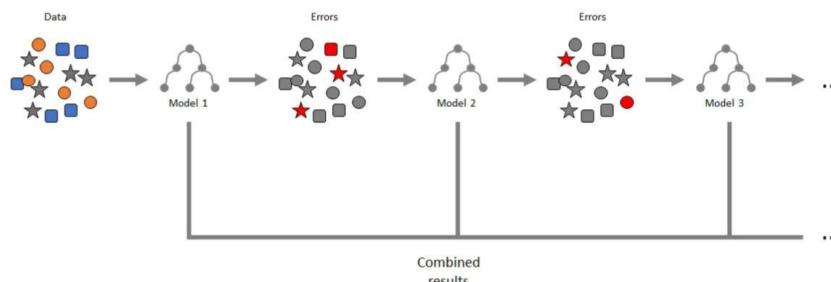
p คือ พารามิเตอร์ที่กำหนด

ถ้า " p " เท่ากับ 1 ระยะทางเท่ากับ Manhattan Distance และ

ถ้า " p " เท่ากับ 2 ระยะทางเท่ากับ Euclidean Distance

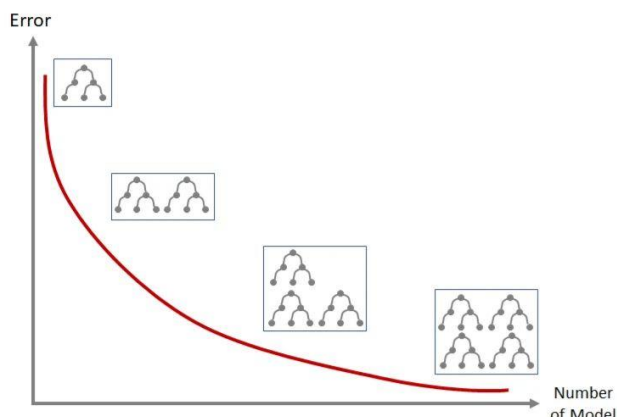
2.1.4 Gradient Boosting (GB)

Gradient Boosting (GB) เป็นการเรียนรู้ของเครื่องแบบ Ensemble Machine Learning (Dhuha T. et al., 2023) แบบ Supervised Learning ทั้ง Regression และ Classification โดยอาศัยหลักการ Decision Tree คือการสร้างต้นไม้หลายๆต้นคล้าย Random Forest แต่การสร้างต้นไม้ใหม่นั้นจะอาศัยการให้น้ำหนักผลการทำนายที่ผิดพลาดจากผลลัพธ์ก่อนหน้ามากกว่าผลการทำนายที่ถูก โดยการทำ 1. การคำนวณหาความผิดพลาดของรอบก่อน 2. การคำนวณหาค่าน้ำหนักของการปรับน้ำหนักของต้นไม้ ซึ่งจะทำให้แบบจำลองมีความแม่นยำในการทำนายมากขึ้น ดังภาพประกอบ 12 และ 13



ภาพประกอบ 12 แสดงโครงสร้าง Gradient Boosting Decision Tree

ที่มา : <https://medium.com/kbtg-life/tree-based-algorithms-%E0%B9%81%E0%B8%9A%E0%B8%9A-high-level-4058e909e0c5>



ภาพประกอบ 13 แสดงแนวโน้มความผิดพลาดกับแบบจำลองที่เพิ่มขึ้น

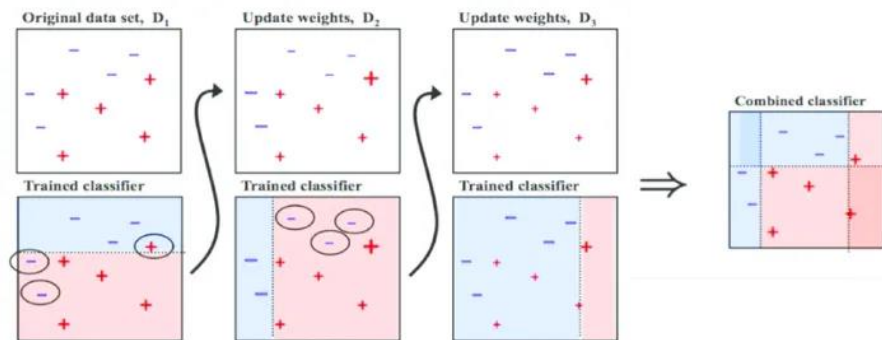
ที่มา : <https://medium.com/kbtg-life/tree-based-algorithms-%E0%B9%81%E0%B8%9A%E0%B8%9A-high-level-4058e909e0c5>

2.1.5 Extreme Gradient Boosting (XGB)

Extreme Gradient Boosting (XGB) หลักการคล้ายกับ Gradient Boosting (GB) เพราะพัฒนามาจาก Gradient Boosting (GB) โดย Extreme Gradient Boosting (XGB) มีการทำ Regularization เพิ่มในส่วนของการปรับน้ำหนักของต้นไม้ เพื่อปรับโครงสร้างต้นไม้และค่าถ่วงน้ำหนักให้การทำนายผิดพลาดในรอบก่อนถูกต้องมากขึ้น จึงช่วยแก้ปัญหา Overfitting ใน Gradient Boosting (GB) และทำให้การเรียนรู้ของแบบจำลองรวดเร็วกว่า Gradient Boosting (GB)

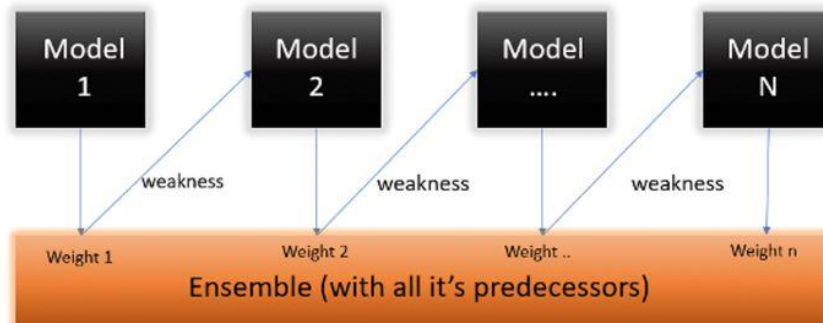
2.1.6 Adaptive Boosting (AdaBoost)

Adaptive Boosting เป็นการเรียนรู้แบบ sequential ensemble method ที่ไม่ซับซ้อน เช่น decision trees (หนึ่งระดับ) ที่เป็นการรวมแบบจำลอง Weak Learner หลายๆ Instance มาฝึกฝนต่อกันเป็นลำดับ เพื่อสร้างแบบจำลองที่ Stronger Learner ที่มีประสิทธิภาพมากขึ้น ใช้หลักการ Reweight ที่ให้ความสำคัญกับจุดข้อมูลที่ทำนายพลาด (misclassified) เพื่อให้ทำนายถูกต้องในครั้งถัดไป การฝึกฝนแต่ละรอบจะมีการกำหนดน้ำหนัก (weight) ของแต่ละชุดข้อมูลและเรียนรู้จากค่าเหล่านั้น ดังภาพประกอบที่ 14 และ 15



ภาพประกอบ 14 แสดงโครงสร้าง Adaptive Boosting (AdaBoost)

ที่มา : <https://ranasinghiitkgp.medium.com/mathematics-behind-random-forest-and-XGBOOSToostooost-ea8596657275>

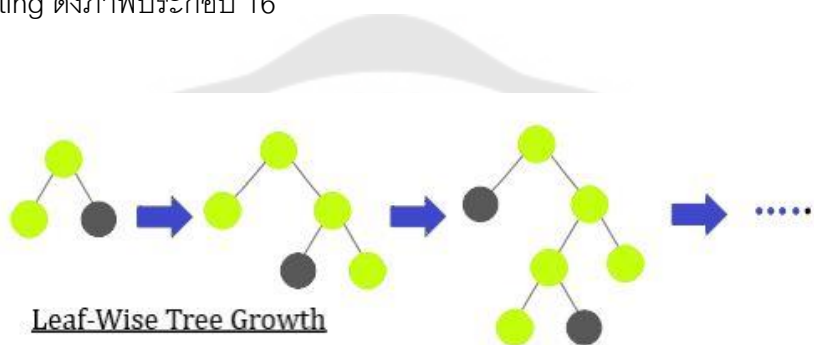


ภาพประกอบ 15 แสดงโครงสร้าง Adaptive Boosting (AdaBoost)

ที่มา : <https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm>

2.1.7 Light Gradient Boosting Machine (LGBM)

ในส่วน Light Gradient Boosting Machine หรือ LGBM เป็น Gradient Boosting Algorithm ที่อาศัยจาก Weak Learners หลายๆแบบจำลองเพื่อสร้างแบบจำลอง Strong Learner เป็นการเรียนรู้แบบ Boosting โดยที่ LGBM ถูกพัฒนาขึ้นโดย Microsoft เน้นการทำงานอย่างรวดเร็วและมีประสิทธิภาพสูงในการสร้างและปรับปรุงแบบจำลอง Gradient Boosting โดยใช้เทคนิค Leaf-Wise Tree Growth เป็นการเจริญต้นไม้รูปแบบใบ คือการเจริญต้นไม้แบบโหนดต่อโหนด (ใบต่อใบ) แทนการเจริญต้นไม้ระดับต่อระดับ และการอัปเดตแบบขนานในแต่ละรอบของ Boosting ดังภาพประกอบ 16



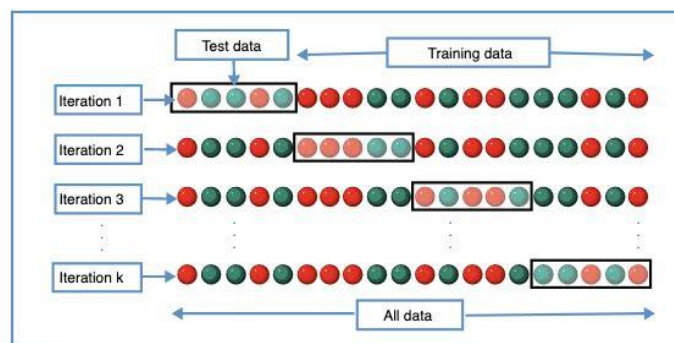
ภาพประกอบ 16 แสดงโครงสร้างแบบจำลอง LGBM ด้วยเทคนิค Leaf-Wise Tree Growth

ที่มา : <https://www.geeksforgeeks.org/lightgbm-light-gradient-boosting-machine>

2.1.8 การแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง (Continuous Learning or Forward-Chaining Split) และ แบบหน้าต่างบานเลื่อน (Sliding Window Split)

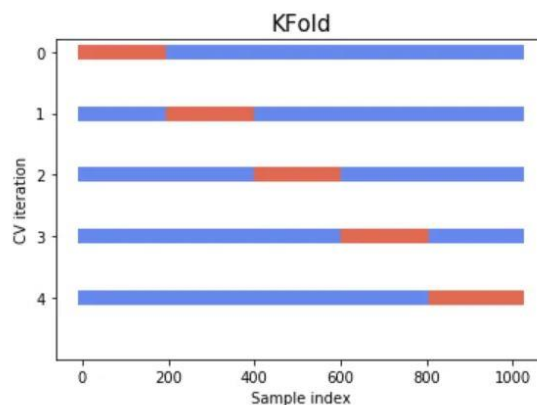
การวิเคราะห์แบบอนุกรมเวลา (Time Series Analysis) เป็นการทำนายข้อมูลที่มีความสัมพันธ์แบบเรียงลำดับ (Sequential data) และการประเมินประสิทธิภาพของแบบจำลองจะสนใจในข้อมูลที่ไม่ได้ใช้ในการฝึกฝน จะเรียกข้อมูลนี้ว่า Unseen Data หรือ Out-of-Sample Data (Tashman, 2000) จะใช้เทคนิค K fold cross validation ดังภาพประกอบที่ 17 และ 18 ในการแบ่งข้อมูลออกเป็นส่วนๆ ในแต่ละส่วนจะกระจายให้ข้อมูลเท่าๆกัน โดยการแบ่งข้อมูลจะ

แบ่งเป็นข้อมูลการฝึกฝน (Train set) และข้อมูลการทดสอบ (Test set) ในการตรวจสอบความถูกต้องของข้อมูล (Validation data) จะทำหลังจากการฝึกฝนแล้วเสมอ โดยใช้วิธีการ คือ Forward-Chaining cross validation และ Sliding Window cross validation ดังภาพประกอบที่ 19



ภาพประกอบ 17 แสดง K fold cross validation

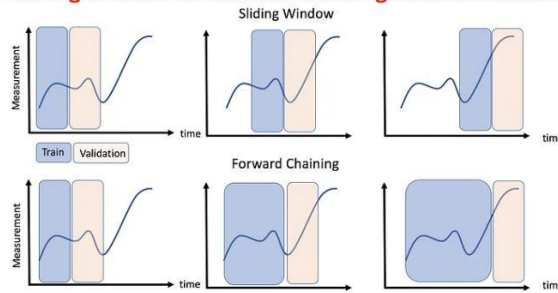
ที่มา : <https://lengyi.medium.com/cross-validation-time-series-model-1cca8424e442>



ภาพประกอบ 18 แสดง K fold ในแต่ละรอบการวนซ้ำ (Iteration)

ที่มา : <https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4>

Sliding window vs. forward chaining cross validation

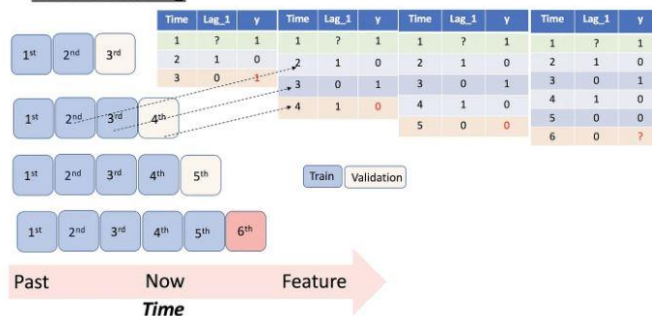


ภาพประกอบ 19 แสดง Continuous Learning หรือ Forward-Chaining Split และ Sliding Window Split

ที่มา : <https://medium.com/@pradip.samuel/cross-validation-in-time-series-model-b07fbba65db7>

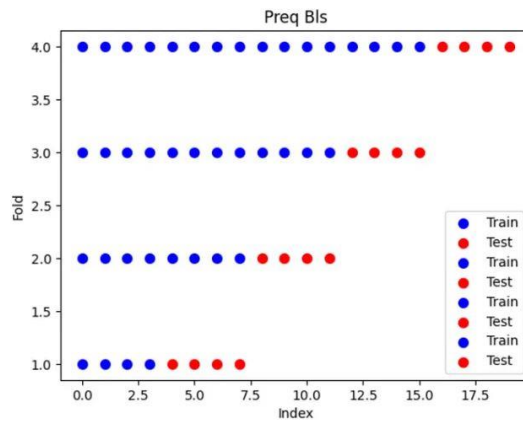
Continuous Learning หรือ Forward-Chaining Split หรือ Rolling-origin เป็นแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง จะกำหนดค่าเริ่มต้นที่เดิมเสมอในการแบ่งข้อมูลของส่วนการทดสอบจะแบ่งเท่าๆกัน และส่วนการฝึกฝนข้อมูลจะเพิ่มขึ้นเรื่อยๆ ดังภาพประกอบที่ 20 และ 21 จากภาพตัวอย่างจะแบ่งการฝึกฝน (Training) เป็น 4 ครั้ง และทำการทดสอบ (Testing) ในแต่ละการวนซ้ำ (Iteration) ให้มีขนาดเท่าๆกัน และค่อยๆเพิ่มข้อมูลฝึกฝนที่เพิ่มขึ้นเรื่อยๆ

Forward Chaining



ภาพประกอบ 20 แสดง Continuous Learning หรือ Forward-Chaining Split

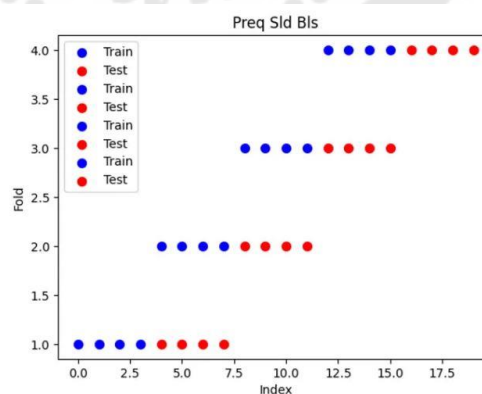
ที่มา : <https://python.plainenglish.io/cross-validation-techniques-for-time-series-data-d1ad7a3a680b>



ภาพประกอบ 21 แสดง Continuous Learning หรือ Forward-Chaining Split

ที่มา : <https://python.plainenglish.io/cross-validation-techniques-for-time-series-data-d1ad7a3a680b>

Sliding Window Split เป็นการแบ่งข้อมูลการเรียนรู้แบบหน้าต่างบ้านเลื่อนจุดเริ่มต้นการฝึกฝนในแต่ละรอบจะไม่เหมือนเดิม จะขยับจุดการเริ่มต้นการฝึกฝนไปเรื่อยๆ เนื่องจากจุดเริ่มต้นการฝึกฝนในรอบถัดไปจะเป็นจุดทดสอบในรอบก่อนหน้า ดังภาพประกอบที่ 22



ภาพประกอบ 22 แสดง Sliding Window Split

ที่มา : <https://python.plainenglish.io/cross-validation-techniques-for-time-series-data-d1ad7a3a680b>

2.1.9 วิธีการประเมินผลแบบจำลอง (Model Evaluation)

ในการวัดค่าประสิทธิภาพแบบจำลองทาง Regression จะมีการวัดด้วยค่าต่างๆ ดังนี้ Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) , R-Squared (R^2)

การวัดประสิทธิภาพแบบจำลองด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) โดยค่า MAE จะมีความอ่อนไหวต่อ Outlier น้อย โดยหาจาก ดังสมการที่ (9)

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y_i| \quad (9)$$

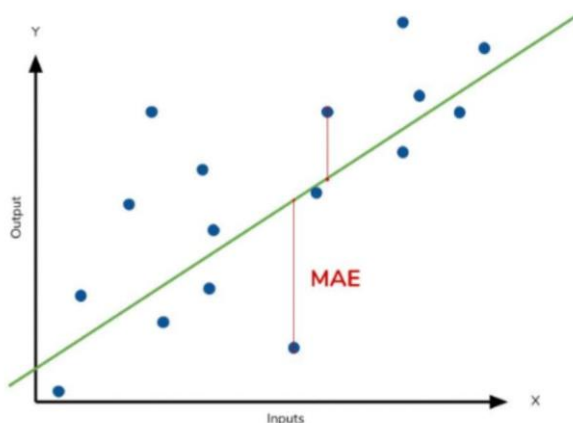
โดยที่

\hat{y} คือ ค่าการทำนาย (prediction value)

y_i คือ ค่าจริง (actual value)

n คือ จำนวนข้อมูลทั้งหมด (Observation)

ค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) (Wang & Lu, 2018) เหมาะกับการวัดความผิดพลาดของแบบจำลองโดยรวม และ เปรียบเทียบแบบจำลองได้อย่างรวดเร็ว โดยยิ่ง MAE มีค่าความผิดพลาดน้อยจะแสดงถึงค่าแบบจำลองมีประสิทธิภาพที่ดี ดังภาพประกอบที่ 23



ภาพประกอบ 23 แสดง Mean Absolute Error (MAE)

ที่มา : <https://hackernoon.com/my-notes-on-mae-vs-mse-error-metrics>

การวัดประสิทธิภาพแบบจำลองด้วยค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2) [https://scikit-learn.org/stable/modules/model_evaluation.html] เป็นการวัดความเหมาะสมของแบบจำลองในบริบทของ Regression เพื่อหาความเหมาะสมกับข้อมูลจริง ดังนั้นจึงเป็นความสำคัญทางค่าสถิติของแบบจำลองที่ใช้ในการทำนายผลหรือการทดสอบทางสมมติฐาน โดยหาจาก ดังสมการที่ (10)

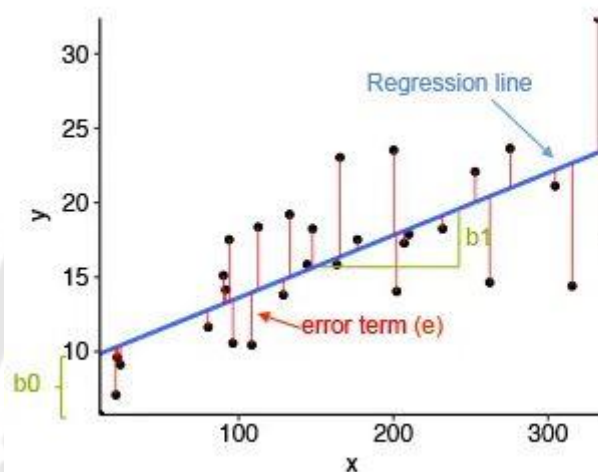
$$\begin{aligned}
 R^2 &= 1 - \frac{\text{Sum Squared Regression (SSR)}}{\text{Total Sum of Squared (SST)}} \\
 &= 1 - \frac{(\sum (y_i - \hat{y}_i)^2)}{(\sum (y_i - \bar{y})^2)} \quad (10)
 \end{aligned}$$

โดยที่

\hat{y} คือ ค่าการทำนาย

\bar{y} คือ ค่าเฉลี่ยของ y

โดยค่า Coefficient of Determination หรือ R^2 (R-Squared) จะมีค่า 0 ถึง 1 โดยค่ายิ่งเข้าใกล้ 1 จะยิ่งดี โดยปกติทั่วไปค่าควรมีค่ามากกว่า 0.6 จึงจะถือว่าเป็นแบบจำลองที่ดี ดังภาพประกอบ 24 จุดสีดำ คือ ค่าของข้อมูลจริง (actual data) ส่วนเส้นตรงสีแดง คือ ค่าการทำนาย (prediction) ที่ได้จากการฝึกสอน (Train)



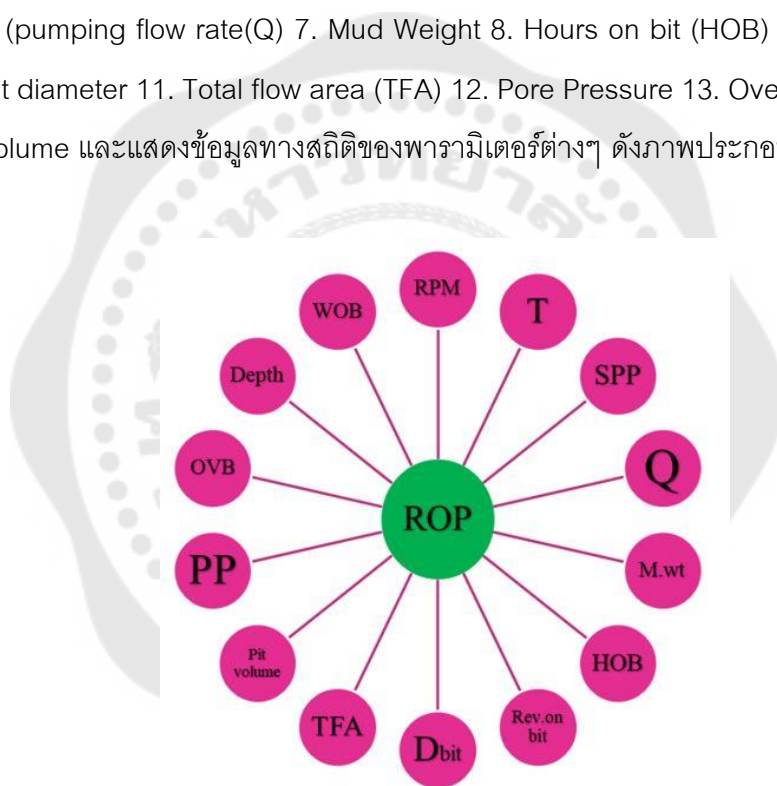
ภาพประกอบ 24 แสดง Linear Regression

ที่ ม า :<https://medium.com/@615162020027/metrics-%E0%B8%9E%E0%B8%B7%E0%B9%89%E0%B8%99%E0%B8%90%E0%B8%B2%E0%B8%99%E0%B8%AA%E0%B8%B3%E0%B8%AB%E0%B8%A3%E0%B8%B1%E0%B8%9A%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%AA%E0%B8%B4%E0%B8%97%E0%B8%98%E0%B8%B4%E0%B8%A0%E0%B8%B2%E0%B8%9E%E0%B8%82%E0%B8%AD%E0%B8%87%E0%B9%82%E0%B8%A1%E0%B9%80%E0%B8%94%E0%B8%A5-machine-learning-c00fcc32fa30>

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 บทความวิจัยเรื่อง Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms (El-Sayed et al., 2023)

บทความวิจัยนี้ได้ทำนายอัตราการเจาะ (Rate Of Penetration , ROP) โดยการเรียนรู้ของเครื่องในการสร้างแบบจำลอง โดยใช้แบบจำลองในการทำนายอัตราการเจาะ (Rate Of Penetration , ROP) ที่มีความแตกต่างกันของพารามิเตอร์ต่างๆ ได้ 1.Depth 2. Weight On Bit (WOB) 3. Revolution per minute (RPM) 4. Torque(T) 5. Standpipe Pressure (SPP) 6. Flow in pump (pumping flow rate(Q) 7. Mud Weight 8. Hours on bit (HOB) 9. Revolutions on bit 10. Bit diameter 11. Total flow area (TFA) 12. Pore Pressure 13. Overburden Pressure 14. Pit Volume และแสดงข้อมูลทางสถิติของพารามิเตอร์ต่างๆ ดังภาพประกอบ 25 และ 26



ภาพประกอบ 25 ปัจจัยที่มีผลต่อการทำนายอัตราการเจาะ

ที่มา : El-Sayed et al.(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

Statistical parameter	ROP (m/hr)	WOB (klb)	RPM (rpm)	T (ft-lb)	SPP (psig)	Q (gpm)	Mwt (ppg)	HOB (hr)	Rev On Bit (krev)	D _{BIT} (in)	TFA in ²	pit volume (bbl)	PP (ppg)	OVB (ppg)
Maximum	25.3	32.8	121	4948.7	3429	1079	10.72	43.85	304	17.50	1.17	785.26	10.32	16.04
Minimum	3.7	0.6	49	1749.7	1645	508	10.20	0.05	0.1	12.25	0.99	396.82	8.85	12.51
Range	21.6	32.2	72	3199	1784	571	0.52	43.80	303.9	5.25	0.17	388.44	1.47	3.53
Mean	12.35	8.45	115.28	2666.47	2908.09	960.78	10.50	19.47	128.27	14.91	1.08	698.69	9.73	14.5
Median	11.60	7.6	119	2623	2952	940	10.50	18.65	122.1	17.50	1.17	695.85	9.81	14.59
Standard Deviation	3.61	4.88	14.51	484.33	291.67	60.73	0.13	11.61	82.94	2.63	0.09	38.50	0.36	1.04

ภาพประกอบ 26 แสดงตารางค่าทางสถิติของพารามิเตอร์ต่างๆ

ที่มา : El-Sayed et al.(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

วัตถุประสงค์ของบทความวิจัยนี้เพื่อการพัฒนาแบบจำลองในการทำนายอัตราการเจาะ (Rate Of Penetration , ROP) เพื่อใช้กับข้อมูลภาคสนามที่เป็นข้อมูล Mud logs และ Petro-Physical ข้อมูลที่ใช้เป็นข้อมูลการเจาะจริงในแนวตรงของหลุมก๊าซ (Vertical gas well) โดยทางบทความวิจัยไม่ได้กล่าวถึงชื่อหลุม แต่ได้ให้ชื่อหลุมว่า Well-X เป็นข้อมูลการเจาะในทะเล บริเวณดินดอนสามเหลี่ยมปากน้ำไนล์ของตะวันออกกลาง (Middle East's offshore Nile Delta) เป็นข้อมูลในส่วน 17.5" และ 12.25" จำนวนข้อมูลทั้งหมด 1771 ข้อมูล อยู่ระหว่างระดับความลึก 1207 เมตร ถึง 2092 เมตร แบบจำลองถูกพัฒนาโดยใช้เทคนิคการเรียนรู้ของเครื่องที่แตกต่างกันสองแบบ คือ 1.เพื่อนบ้านที่ใกล้ที่สุด (KNN) เป็นแบบจำลองที่ทำนายอัตราการเจาะสำหรับชุดข้อมูลการฝึกสอนแสดงค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.94. ชุดข้อมูลทดสอบแสดงค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.84 และ 2.แบบจำลอง Multilayer Perceptron (MLP) ทำนายอัตราการเจาะสำหรับชุดข้อมูลการฝึกสอนค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.98 ชุดข้อมูลทดสอบแสดงค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.79 สรุปได้ว่า MLP มีความแม่นยำดีกว่า KNN เมื่อทำการนำค่าผิดปกติ (outlier) ออกด้วยเปอร์เซ็นต์ 14.7% จะช่วยเพิ่มประสิทธิภาพการทำงานของแบบจำลอง Multilayer Perceptron ทำให้เป็นอัลกอริทึมที่ดีที่สุด ในบรรดาอัลกอริทึมที่ศึกษาทั้งสองซึ่งมีค่าสัมประสิทธิ์การตัดสินใจ (R^2) ของทั้งหมดเท่ากับ 92% โดยทางบทความวิจัยได้สรุปผลและสร้างกราฟเปรียบเทียบของทั้ง 2 แบบจำลอง ดังภาพประกอบ 27 และ 28

Model	Dataset	RMSE	R ²
KNN	Train	0.606917	0.94326
	Test	1.837651	0.84054
	Validation	1.654272	0.85156
	Total	0.923941	0.91485
MLP	Train	0.265653	0.97502
	Test	2.528778	0.79381
	Validation	1.962094	0.83643
	Total	0.911712	0.92034

ภาพประกอบ 27 แสดงตารางผลของแบบจำลอง KNN และ MLP

ที่มา : El-Sayed et al.(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

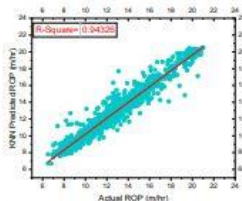


Figure 14. Regression plot of the actual ROP and predicted ROP for training dataset for KNN model.

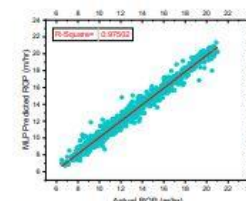


Figure 15. Regression plot of the actual ROP and predicted ROP for training dataset for MLP model.

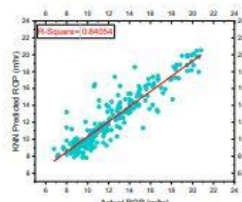


Figure 16. Regression plot of the actual ROP and predicted ROP for test dataset for KNN model.

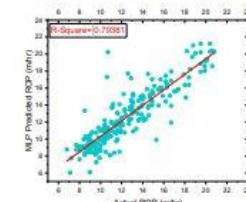


Figure 17. Regression plot of the actual ROP and predicted ROP for test dataset for MLP model.

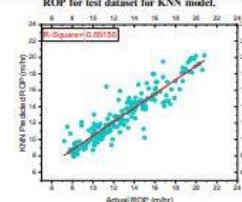


Figure 18. Regression plot of the actual ROP and predicted ROP for validation dataset for KNN model.

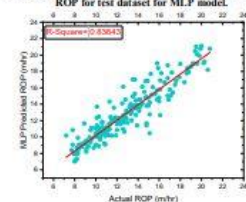
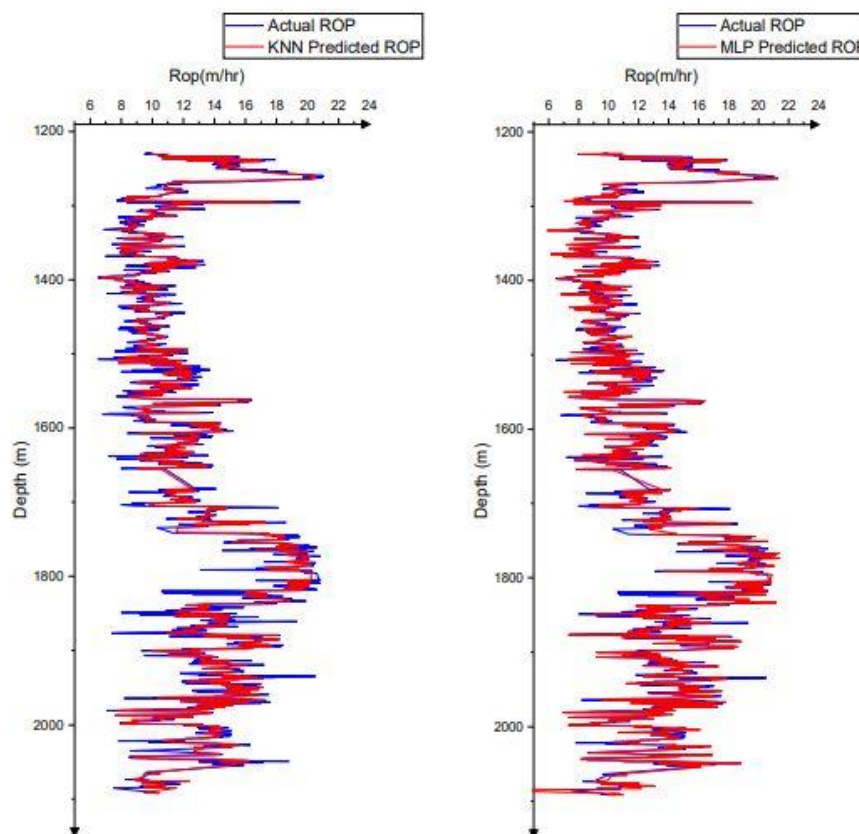


Figure 19. Regression plot of the actual ROP and predicted ROP for validation dataset for MLP model.

ภาพประกอบ 28 แสดงกราฟผลของแบบจำลอง KNN และ MLP

ที่มา : El-Sayed et al.(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

ทั้งนี้ในบทความวิจัยยังได้แสดงกราฟการเปรียบเทียบเทียบของการทำนายอัตราการเจาะกับอัตราการเจาะจริงในช่วงของระดับความลึกต่างๆ ทั้งแบบจำลอง KNN และ MLP ดังภาพประกอบ 29



ภาพประกอบ 29 แสดงการเปรียบเทียบเทียบของการทำนายอัตราการเจาะกับอัตราการเจาะจริงในอดีตในช่วงของระดับความลึกต่างๆในแบบจำลอง KNN และ MLP

ที่มา : El-Sayed et al.(2023). Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

2.2.2 บทความวิจัยเรื่อง Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models (Hegde et al., 2017)

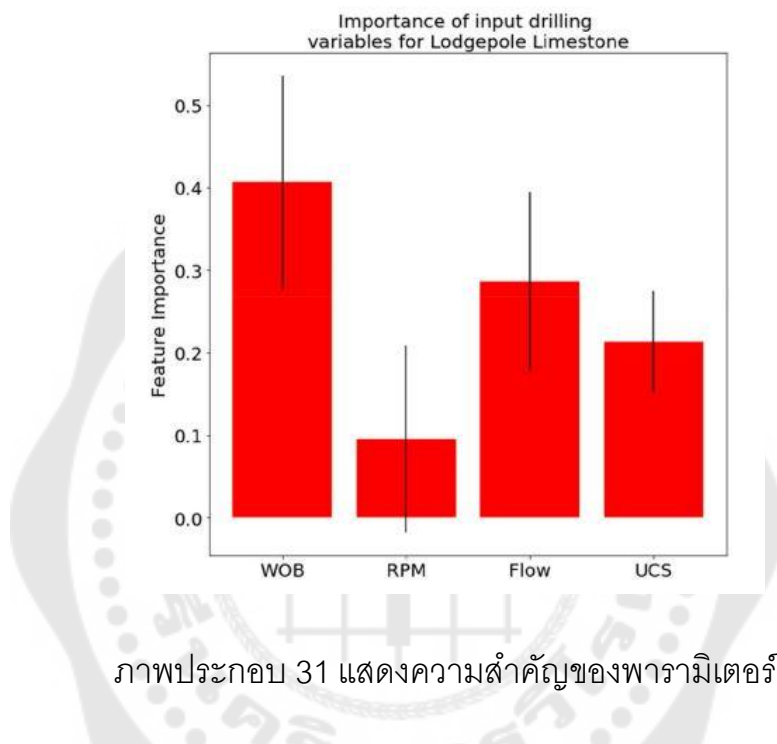
บทความวิจัยนี้ได้ทำนายอัตราการเจาะ (Rate Of Penetration , ROP) โดยทำการเปรียบเทียบแบบจำลองแบบดั้งเดิม (traditional models) ซึ่งเป็นแบบจำลองพื้นฐานทางด้านฟิสิกส์ (Physics-based) แบ่งได้เป็น 3 แบบจำลอง 1. Bingham Model 2. Motahhari Model 3. Hareland Model เปรียบเทียบกับแบบจำลองแบบข้อมูล (Data-driven models) แบ่งได้เป็น 2 แบบจำลอง 1. Linear Regression Model 2. Machine Learning Model (Random Forest) ค่าพารามิเตอร์ที่ใช้ ได้แก่ 1. Weight on bit (WOB) 2.Revolution per minute (RPM) 3.Flow rate เป็นค่าพารามิเตอร์หลักที่เหมือนกัน แต่ในแบบจำลองแบบเดิม (Traditional Models) ทั้ง 3 แบบจำลองยังต้องการค่าพารามิเตอร์เพิ่มอีก ได้แก่ ความแข็งของหิน (The Strength of rock) ซึ่งในแต่ละหมวดหิน (Rock Formations) จะมีค่าความแข็งที่แตกต่างกัน โดยข้อมูลของหมวดหินที่ใช้มาบริษัท Marathon Oil กับข้อมูลหัวเจาะ Smith 616 PDC bit จากการเจาะในแอ่งวิลลิสตันของนอร์ทดาโคตา (Williston basin, North Dakota) เป็นช่วงข้อมูลระดับความลึกจริงในแนวตรง (True vertical depth : TVD) ในระดับความลึก 4871 ฟุต ถึง 9128 ฟุต ซึ่งเป็นข้อมูลใน Bakken shale เป็นหนึ่งในหมวดหินของแอ่งวิลลิสตัน ดังภาพประกอบ 30

SYSTEM	FORMATION OR GROUP		
PERMIAN	Minnekahta		
	Opeche		
	Broom Creek		
PENNSYLVANIAN	Amsden		
	Tyler		
MISSISSIPPIAN	Otter		
	Kibbey		
	Madison Group	Charles Fm.	Poplar interval
			Base last salt
			Ratcliffe interval
	Mission Canyon		
	Lodgepole		
Bakken			
DEVONIAN	Three Forks		

ภาพประกอบ 30 แสดงตารางหมวดหินใน Williston basin, North Dakota

ที่มา : Hegde et al.(2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models

แบบจำลองแบบดั้งเดิมของ Motahhari Model ได้กำหนดให้ในการสมการที่ใช้สร้างแบบจำลองเพื่อหาอัตราการเจาะ (Rate Of Penetration , ROP) มีค่ากำลังอัดแกนเดียว (unconfined rock strength : UCS) ทำให้พารามิเตอร์ที่สำคัญในการทำนาย ROP นั้นจะประกอบด้วย 4 พารามิเตอร์ ดังภาพประกอบ 31



ภาพประกอบ 31 แสดงความสำคัญของพารามิเตอร์

ที่มา : Hegde et al.(2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models

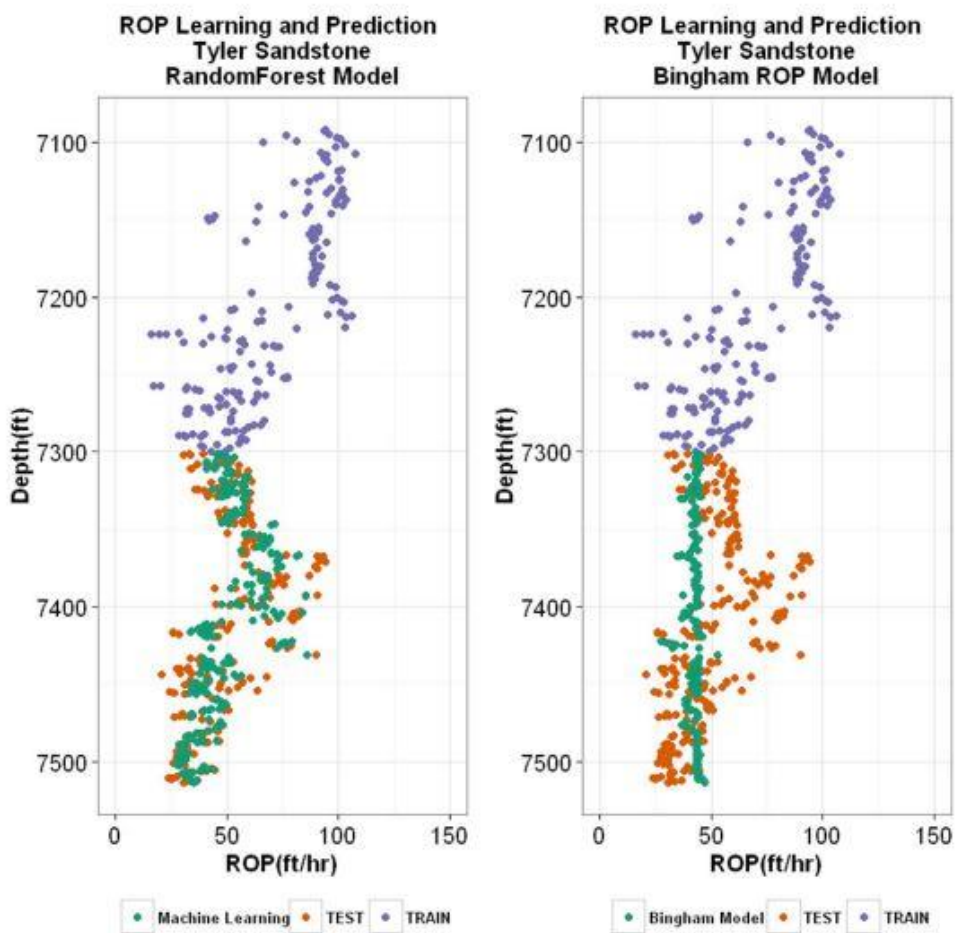
ผลจากการทำนาย ROP แบบจำลองแบบข้อมูล (Data Driven Model) จะให้ค่าความถูกต้องดีกว่าแบบจำลองแบบเดิม (Traditional Model) ดังภาพประกอบ 32

Predictor	R ²	Normalized error	RMSE
Random Forest	0.84	0.12	8.3
Ensemble ROP Model	0.72	0.14	16.21
Linear Regression	0.12	0.35	32
Bingham	0.46	0.26	17.67
Motahhari	0.63	0.23	16
Hareland	0.19	0.36	13.7

ภาพประกอบ 32 แสดงตารางผล R² และ RMSE ของการทำนายอัตราการเจาะในแบบจำลองต่างๆ

ที่มา : Hegde et al.(2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models

เมื่อนำข้อมูลของ Tyler Formation มาทำการ Normalization และ ปรับจูนค่าพารามิเตอร์ (Hyperparameter Tuning) ของแบบจำลอง Random forest และแบบจำลอง Bingham จากนั้นทำการพล็อตกราฟระหว่าง ROP กับ ระดับความลึก (Depth) โดยกำหนดให้ สีฟ้า(Blue)เป็นข้อมูลการทดสอบ (Test data), สีส้ม (Orange) เป็นข้อมูลการฝึกสอน (Training data) และสีเขียว (Green) เป็นข้อมูลการทำนาย ROP แสดงให้เห็นว่าแบบจำลองแบบเรียนรู้ด้วยเครื่อง (Machine Learning) แบบ Random forest ให้ค่า R² = 0.97 และแบบจำลองแบบดั้งเดิม (Traditional model) แบบ Bingham ให้ค่า R² = 0.64 ดังภาพประกอบ 33



ภาพประกอบ 33 แสดงการเปรียบเทียบข้อมูลการทดสอบ (Test data), ข้อมูลการฝึกสอน (Training data) และข้อมูลการทำนายอัตราเจาะ (Rate Of Penetration , ROP) ในแบบจำลอง Random forest และ Bingham

ที่มา : Hegde et al.(2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models

2.2.3 บทความวิจัยเรื่อง Research on the Rate of Penetration Prediction Method Based on Stacking Ensemble Learning (Ren et al., 2023)

บทความวิจัยนี้ได้ทำนายอัตราเจาะ (Rate Of Penetration , ROP) โดยเปรียบเทียบระหว่างแบบจำลอง K-nearest neighbor (KNN) และ Support Vector Machine (SVM) โดยใช้ 1. Stacking Algorithm เป็นการเรียนรู้แบบผสมผสาน (ensemble learning) ที่อาศัยตัวเรียน (learner) หลายตัวในการฝึกฝน โดยแบ่งตัวเรียนออกเป็น 2 ประเภท ได้แก่ primary learner และ meta-learner โดย primary learner ทำหน้าที่ฝึกฝนข้อมูลแยกตามชุดข้อมูล ขณะที่ meta-learner ทำหน้าที่ฝึกฝนข้อมูลที่ได้จากการรวมชุดข้อมูลเข้าด้วยกัน stacking algorithm สามารถดึงจุดเด่นของแบบจำลองการทำนายที่แตกต่างกันมาใช้ ประยุกต์ใช้ primary learner หลายตัวร่วมกันแบบไม่จำกัดชั้น (infinitely superpose) และใช้ meta-learner ปรับแต่งการผสมผสานระหว่าง primary learner ต่างๆ ด้วยเหตุนี้ เมื่อเทียบกับ single prediction model จะเห็นว่า stacking algorithm จึงมี ประสิทธิภาพที่ดีกว่า (better performance) , ความแข็งแกร่ง (stronger robustness) และการสรุป (generalization) ที่เหนือกว่า และ 2. Genetic Algorithm (GA) เป็นอัลกอริทึมทางพันธุกรรมเป็นวิธีการค้นหาแบบสุ่มและการหาค่าเหมาะสมที่มีประสิทธิภาพ โดยอาศัยหลักการทางพันธุกรรมตามธรรมชาติและวิวัฒนาการทางชีววิทยา ในกระบวนการแก้ปัญหา อัลกอริทึมเลียนแบบกลไกการวิวัฒนาการของโลกชีวภาพ จากนั้นจึงได้ผลลัพธ์ที่เหมาะสมที่สุดผ่านการทำงานของขั้นตอนการคัดเลือกโครโมโซม การข้ามพันธุ (crossover) และการกลายพันธุ์ (mutation) หัวใจสำคัญของอัลกอริทึมนี้คือการประเมินคุณค่าของแต่ละตัวเลือก (individuals) ผ่าน fitness function ตัวเลือกที่โดดเด่นจะมีสัดส่วนที่สูงกว่าในประชากรของรุ่นถัดไป เมื่อเปรียบเทียบกับอัลกอริทึมการหาค่าเหมาะสมแบบดั้งเดิม อัลกอริทึมทางพันธุกรรมมีความแข็งแกร่ง (robustness) ที่ดี รวมถึง ลักษณะการค้นหาแบบทั่วโลก (global search) และแนวโน้มสู่การทำงานแบบขนาน (parallelism) ข้อมูลที่ใช้มาจากแหล่งน้ำมัน Xinjiang (Xinjiang) ประเทศจีน (China) โดยการทำนายด้วยแบบจำลอง KNN , SVM และ stacking models ให้ผลการทำนาย 63.2% , 73.7% และ 81.6% ตามลำดับ ขณะที่ค่า Parameters ของแต่ละแบบจำลองที่ใช้ Genetic Algorithm ใน KNN, SVM และ stacking models ให้ผลการทำนาย 73.7% , 78.9% และ 92.5% ตามลำดับ ซึ่งผลการทำนายแสดงให้เห็นว่าความแม่นยำของการทำนาย ROP สูงถึง 92.5% และมีประสิทธิภาพคงที่ ซึ่งสามารถใช้เป็นข้อมูลอ้างอิงสำหรับการปรับแต่งพารามิเตอร์ในการเจาะแหล่งน้ำมันแห่งนี้ และเป็นแนวทางในการยกระดับประสิทธิภาพของการดำเนินงานเจาะ ดังภาพประกอบที่ 34

TABLE 5: Default model prediction results.

Model	Accuracy of classification
KNN	63.2%
SVM	73.7%
Stacking	81.6%

TABLE 9: Model prediction results optimized by genetic algorithm.

Model	Accuracy of classification
KNN	73.7%
SVM	78.9%
Stacking	92.5%

ภาพประกอบ 34 แสดงประสิทธิภาพของแบบจำลองด้วย Default model และ Genetic Algorithms

ที่มา : Ren et al.(2023).Research on the Rate of Penetration Prediction Method Based on Stacking Ensemble Learning

2.2.4 บทความวิจัยเรื่อง Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models (Soares & Gray, 2019)


บทความวิจัยนี้ได้แสดงให้เห็นว่าการเรียนรู้ของเครื่อง (Machine Learning) สามารถทำนายอัตราในการขุดเจาะ (ROP) ได้อย่างแม่นยำกว่าแบบจำลองเชิงวิเคราะห์ (Analytical models) โดยใช้ข้อมูลพื้นผิว (Surface data) เดียวกัน ค่าพารามิเตอร์ที่ส่งผลกระทบต่อค่า ROP ได้แก่ 1. Weight on bit (WOB) 2. Rotation Speed (RPM) 3. Drilling fluid rate สำหรับชุดข้อมูล Williston Basin ที่วิเคราะห์ พบว่าแบบจำลอง Bourgoyne และ Young ที่ปรับเปลี่ยนตามที่เสนอ ในบทความวิจัยนี้มีประสิทธิภาพดีที่สุดในบรรดาแบบจำลอง ROP เชิงวิเคราะห์ ดังภาพประกอบที่ 35 ในขณะที่แบบจำลอง Random Forest โดยรวมมีข้อผิดพลาดน้อยที่สุด สามารถจัดการเรียนรู้แบบใหม่ได้ว่าแบบจำลองการเรียนรู้ของเครื่องสามารถลดข้อผิดพลาดของการทดสอบได้อย่างมีประสิทธิภาพมากขึ้นเมื่อมีข้อมูลการฝึกอบรมมากขึ้น ขณะที่ปริมาณข้อมูลที่จำเป็นสำหรับวัดประสิทธิภาพของการเรียนรู้ของเครื่องน้อยกว่าที่คาดไว้เนื่องจากแบบจำลองการเรียนรู้ของเครื่องแสดงข้อผิดพลาดน้อยกว่าแบบจำลองเชิงวิเคราะห์ในช่วงเวลาการปรับแต่งที่มีข้อมูลเพียงสิบจุด สำหรับการฝึกอบรม การฝึกฝนแบบจำลองหลายๆ ส่งผลให้ประสิทธิภาพของแบบจำลองดีขึ้นแสดงให้เห็นถึงคุณค่าของการเรียนรู้ของแบบจำลองอย่างต่อเนื่องแบบเรียลไทม์ การตรวจสอบข้าม (Cross Validation) ให้ผลลัพธ์ที่น่าพอใจในฐานะเป็นเทคนิคการเลือกแบบจำลองที่มีความแม่นยำในการทำนาย ROP สูงแบบเรียลไทม์ ขั้นตอนการเลือกแบบจำลองนี้มีความสำคัญ

เนื่องจากไม่มีข้อมูลการเจาะในพื้นที่ เมื่อเจาะหลุมแบบสำรวจ การควบคุมพารามิเตอร์การเจาะแบบ close-loop ที่เชื่อถือได้นั้นอาศัยการสร้างแบบจำลอง ROP ที่แม่นยำ ดังนั้นผลในบทความวิจัยนี้จะส่งผลกระทบต่อการใช้การขุดเจาะแบบอัตโนมัติ

Table 1
Drilling variables for four analytical ROP models divided into data types.

Analytical ROP Model	Bingham (1964)	Bourgoyne & Young (1974)	Hareland and Rempersad (1994)	Motahhari et al. (2010)
Operational Variables				
Weight-on-Bit, WOB	✓	✓	✓	✓
Rotary Speed, RPM	✓	✓	✓	✓
Flow Rate, q		✓		
Well Depth, D		✓		
Equivalent Circulating Density, ρ_c		✓		
Bit Wear, h or W_f		✓	✓	✓
Rock Properties				
Pore Pressure Gradient, g_p		✓		
Unconfined Compressive Strength, σ_c			✓	
Confined Compressive Strength, S				✓
Bit Properties				
Bit Diameter, d_b	✓	✓	✓	✓
Bit Nozzle Diameter, d_n		✓		
PDC Cutters Design Properties				
Number of Cutters, N_c			✓	✓
Cutter Diameter, d_c			✓	
Cutter Siderake Angle, α			✓	
Cutter Backrake Angle, θ			✓	
Drilling Fluid Properties				
Mud Density, ρ		✓		
Apparent Viscosity at $10,000 \text{ sec}^{-1}$, μ		✓		

 surface measurements available in real-time

 uncertain data availability in real-time

 constant during a bit run

ภาพประกอบ 35 แสดงการเปรียบเทียบปัจจัยการขุดเจาะใน 4 แบบจำลองเชิงวิเคราะห์ ROP

ที่มา : Soares & Gray.(2019). Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models

2.2.5 บทความวิจัยเรื่อง Reference dataset for rate of penetration benchmarking

(Tunkiel et al., 2021)

บทความวิจัยนี้ใช้ชุดข้อมูลเผยแพร่จากบริษัท Equinor ปี 2018 ชุดข้อมูลชื่อ Volve dataset ข้อมูลหลุมทั้งหมด 7 หลุม พารามิเตอร์ที่ใช้ ได้แก่ 1.Measured Depth [m] 2.Weight on Bit [kkgf] 3.Average Standpipe Pressure [kPa] 4.Average Surface Torque [kN.m] 5.Rate of Penetration [m/h] 6.Average Rotary Speed [rpm] 7.Mud Flow In [L/min] 8.Mud Density In [g/cm³] 9.Diameter [mm] 10.Average Hookload [kg] 11.Hole Depth (TVD) [m] 12.USROP Gamma [gAPI] โดยเปรียบเทียบแบบจำลองทั้งหมด 6 แบบจำลอง มี 1 แบบจำลองแบบเดิมทางฟิสิกส์ (physics-based model) ได้แก่ Bingham แบบจำลองนี้ได้รับการยอมรับและใช้งานโดยนักวิจัยมากที่สุด เนื่องจากมีความชัดเจนในการนำไปใช้ และอีก 5 แบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) ได้แก่ Gradient Boosting, Random Forest, AdaBoost, K-Nearest Neighbors และ XGB ใช้วิธีการฝึกฝนและการทดสอบแบบ TimeSeriesSplit โดยเทคนิคที่ใช้ คือ Continuous Learning, All for One และ One for All ดังภาพประกอบที่ 36 และวัดค่าประสิทธิภาพแบบจำลองด้วย Mean Absolute Error (MAE) และ Weighted Mean Absolute Percentage Error (WMAPE) ในการรวมหลุมทั้งหมดเข้าด้วยกัน และ Train-Test Split อัตราส่วน 9:1 ได้ผลลัพธ์ Gradient Boosting ที่ไม่ได้ปรับแต่ง MAE เท่ากับ 4.75 และ R2 เท่ากับ 0.82 และ แบบจำลอง Bingham ใน Continuous Learning ที่ใช้ชุดข้อมูล USROP ให้ค่า WMAPE อยู่ระหว่าง 20% ถึง 103% ผลวิจัยสรุปว่า เมื่อแบ่งข้อมูลตามวิธีที่สมจริงเหมาะสำหรับการใช้งานภาคสนาม แบบจำลองการเรียนรู้ของเครื่องแบบมาตรฐานจะมีประสิทธิภาพต่ำกว่างานวิจัยก่อนหน้านี้ ผลวิจัยที่นำเสนอนี้ช่วยชี้ให้เห็นถึงประสิทธิภาพของวิธีการที่ใช้ข้อมูล ซึ่งขึ้นอยู่กับวิธีการและหลุมเจาะเฉพาะเจาะจง อาจจะเหนือกว่าหรือด้อยกว่าแบบจำลองทางฟิสิกส์แบบคลาสสิกอย่างมากในบางครั้ง ยิ่งยืนยันถึงความจำเป็นสำหรับชุดข้อมูลอ้างอิงขนาดใหญ่ ซึ่งช่วยให้สามารถประเมินความแม่นยำของโมเดลที่พัฒนาขึ้นได้อย่างแข็งแกร่ง

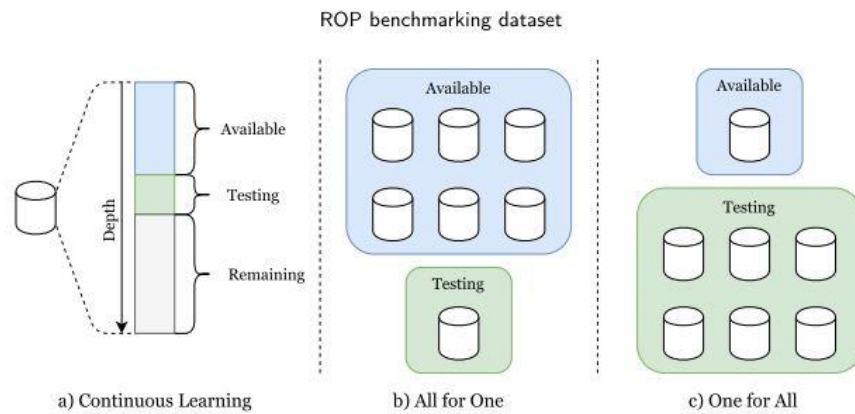


Fig. 3. Proposed scenarios for model validation.

ภาพประกอบ 36 แสดงผลการแบ่งข้อมูลเพื่อการฝึกฝนและการทดสอบ

ที่มา : Tunkiel et al.(2021).Reference dataset for rate of penetration benchmarking

บทที่ 3

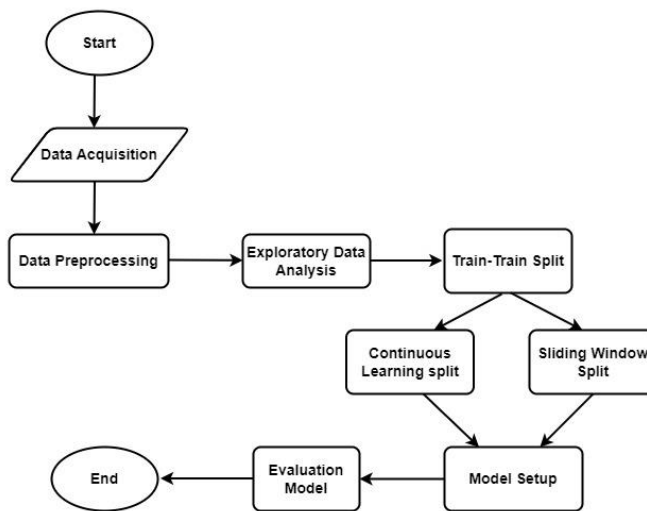
การดำเนินงานวิจัย

การศึกษางานวิจัยนี้เป็นการวิจัยเพื่อการทำนายหาค่าความแม่นยำของอัตราการเจาะเทียบกับค่าอัตราการเจาะจริงของหลุมปิโตรเลียม เพื่อช่วยในการประเมินวางแผนงานระยะเวลาการเจาะและประเมินค่าใช้จ่ายในการขุดเจาะของหลุมปิโตรเลียม รวมถึงยังเป็นแนวทางให้วิศวกรขุดเจาะเพื่อใช้ประกอบการตัดสินใจในหน่วยงานขณะที่มีการดำเนินการขุดเจาะอยู่ นั้น โดยการวิจัยนี้ศึกษาวิเคราะห์เปรียบเทียบแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งประกอบด้วย Random Forest (RF), K-Nearest Neighbor (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost) และ Light Gradient Boosting Machine (LGBM)

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอน ดังต่อไปนี้

- 3.1 กระบวนการทำงานของแบบจำลอง (Model Overview Workflow)
- 3.2 การเตรียมข้อมูล (Data Preparation)
 - 3.2.1 ที่มาของแหล่งข้อมูล (Data Source)
 - 3.2.2 การได้มาของข้อมูล (Data Acquisition)
 - 3.2.3 การทำความสะอาดและการเตรียมข้อมูล (Data Cleaning and Preprocessing)
 - 3.2.4 การเลือกค่าพารามิเตอร์ที่เกี่ยวข้อง (Selection Parameters)
- 3.3 การสำรวจข้อมูลคุณลักษณะเบื้องต้น (Exploratory Data Analysis)
 - 3.3.1 ข้อมูลที่ผิดปกติ (Outliers)
 - 3.3.2 การหาความสัมพันธ์ระหว่างตัวแปร (Parameters Correlation)
 - 3.3.3 การอธิบายค่าตัวแปรต่างๆ (Explain Feature)
- 3.4 การตั้งค่าแบบจำลองและการฝึกฝน (Model Setup and Training)
 - 3.4.1 การแบ่งชุดข้อมูลการฝึกฝนและการทดสอบ (Train-Test Split)
 - 3.4.2 การตั้งค่าแบบจำลอง (Model setup)
- 3.5 วิธีการประเมินผลแบบจำลอง (Model Evaluation)

3.1 กระบวนการทำงานของแบบจำลอง (Model Overview Workflow)



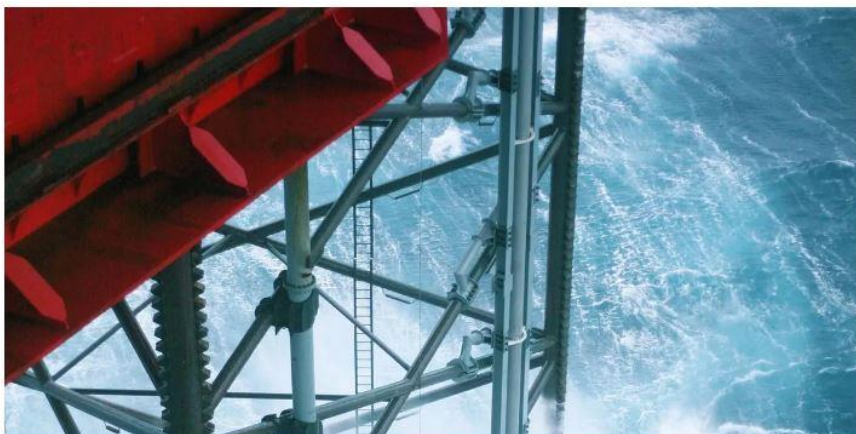
ภาพประกอบ 37 แสดงกระบวนการทำงานของแบบจำลอง (Model Overview Workflow)

3.2. การเตรียมข้อมูล (Data Preparation)

3.2.1 ที่มาของแหล่งข้อมูล (Data Source)

ข้อมูลที่ใช้ในการดำเนินงานวิจัยนี้ นำมาจากบริษัท Equinor ซึ่งเป็นรัฐบริษัทพลังงานแห่งชาติของประเทศนอร์เวย์ โดยข้อมูลนี้ได้เผยแพร่จากทางบริษัทเมื่อปี 2018 เพื่อการศึกษาและวิจัยเท่านั้น เป็นชุดข้อมูลจริงที่สามารถดาวน์โหลดชุดข้อมูลได้ที่ <https://www.equinor.com/energy/volve-data-sharing> ชุดข้อมูลได้มาจากแหล่งไหล่ทวีปของนอร์เวย์ (Norwegian Continental shelf : NCS) ซึ่งอยู่ใน Volve field มีพื้นที่ตั้งอยู่ใจกลางทะเลเหนือ (Central part of North Sea) ห่างจาก Stavanger ประมาณ 201 กิโลเมตร (ประมาณ 125 ไมล์) และ 5 กิโลเมตร ทางด้านเหนือของ Sleipner ตะวันออก ดังภาพประกอบ 38 ถึง 41

Volve field data set



The Maersk Inspire platform at the Volve field
Photo: Calin Dobinson

Home > Energy > Data sharing > Volve data sharing

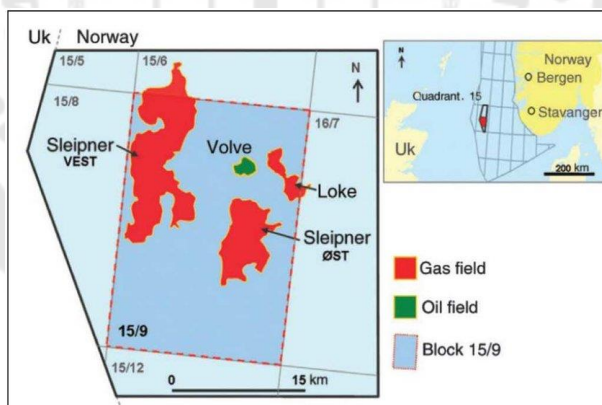
ภาพประกอบ 38 แสดงหน้า website บริษัท Equinor ที่สามารถดาวน์โหลดข้อมูล
ได้

ที่มา : <https://www.equinor.com/energy/volve-data-sharing>



ภาพประกอบ 40 แสดงจุดพื้นที่ของ Volve filed ใจกลางทะเลเหนือ

ที่มา : Wang et al. (2021). Ensemble Machine Learning Assisted Reservoir Characterization Using Field Production Data—An Offshore Field Case Study. Energies.



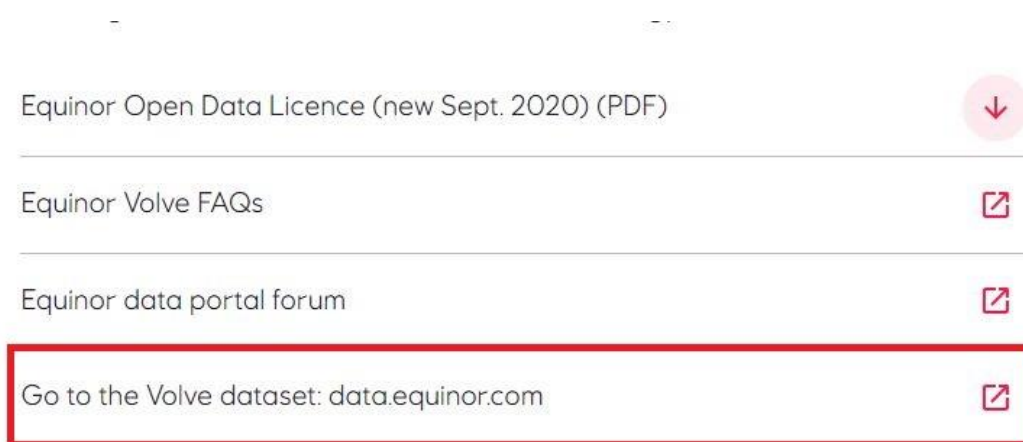
ภาพประกอบ 41 แสดงแหล่งพื้นที่ของ Volve filed

ที่มา : Wang et al. (2021). Ensemble Machine Learning Assisted Reservoir Characterization Using Field Production Data—An Offshore Field Case Study. Energies.

3.2.2 การได้มาของข้อมูล (Data Acquisition)

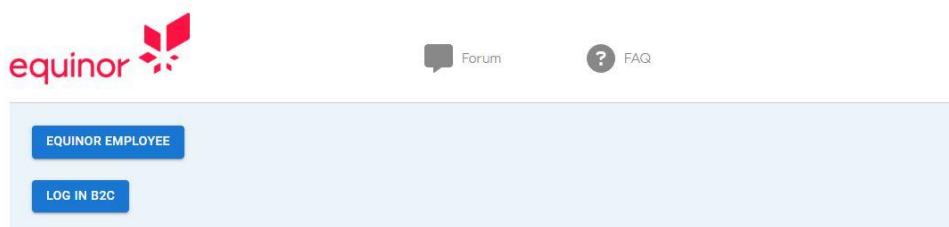
การดึงชุดข้อมูลของ Volve จะประกอบไปด้วย 2 ส่วน ดังนี้

1.เข้ามาที่ website <https://www.equinor.com/energy/volve-data-sharing> เพื่อทำการ Log in โดยเลือก Log in B2C (กรณีที่ไม่ใช่พนักงานบริษัท) โดยชุดข้อมูลที่ใช้มีชื่อว่า Volve Data Village เป็นข้อมูล Subsurface และ Production เมื่อเข้าไปที่ชุดข้อมูลนี้จะสามารถรับ Azure signature URL ดังภาพประกอบ 42 ถึง 45



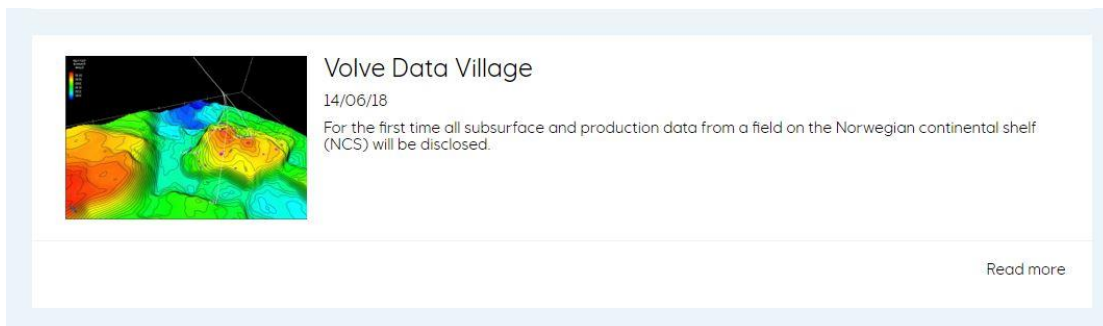
ภาพประกอบ 42 แสดงการเข้าถึงการดาวน์โหลดชุดข้อมูล

ที่มา : <https://www.equinor.com/energy/volve-data-sharing>



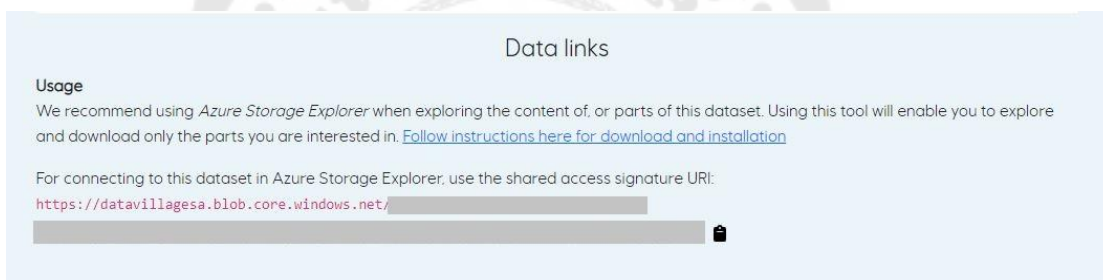
ภาพประกอบ 43 แสดงหน้าการ Login B2C

ที่มา : <https://www.equinor.com/energy/volve-data-sharing>



ภาพประกอบ 44 แสดงข้อมูล Volve Data Village

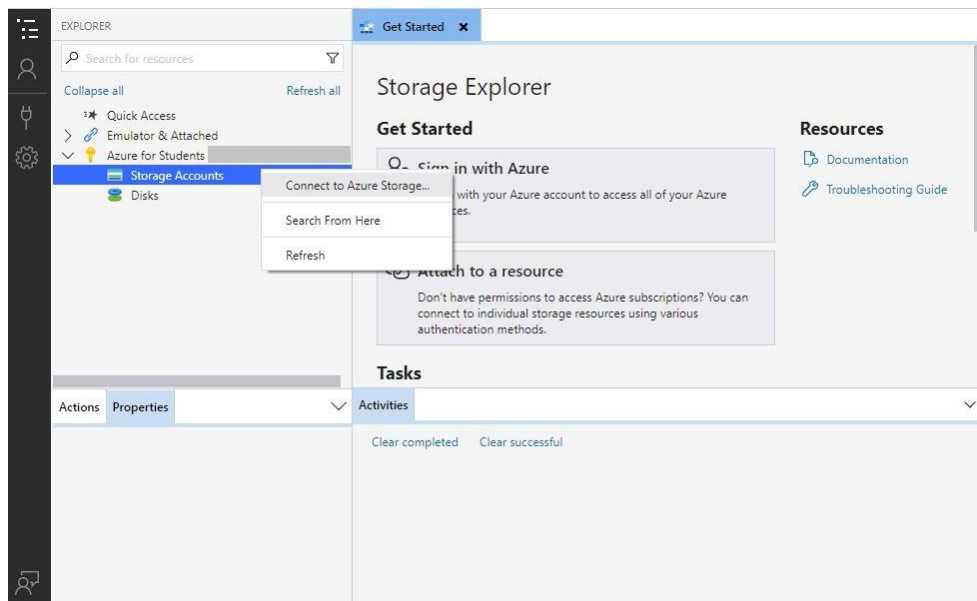
ที่มา : <https://www.equinor.com/energy/volve-data-sharing>



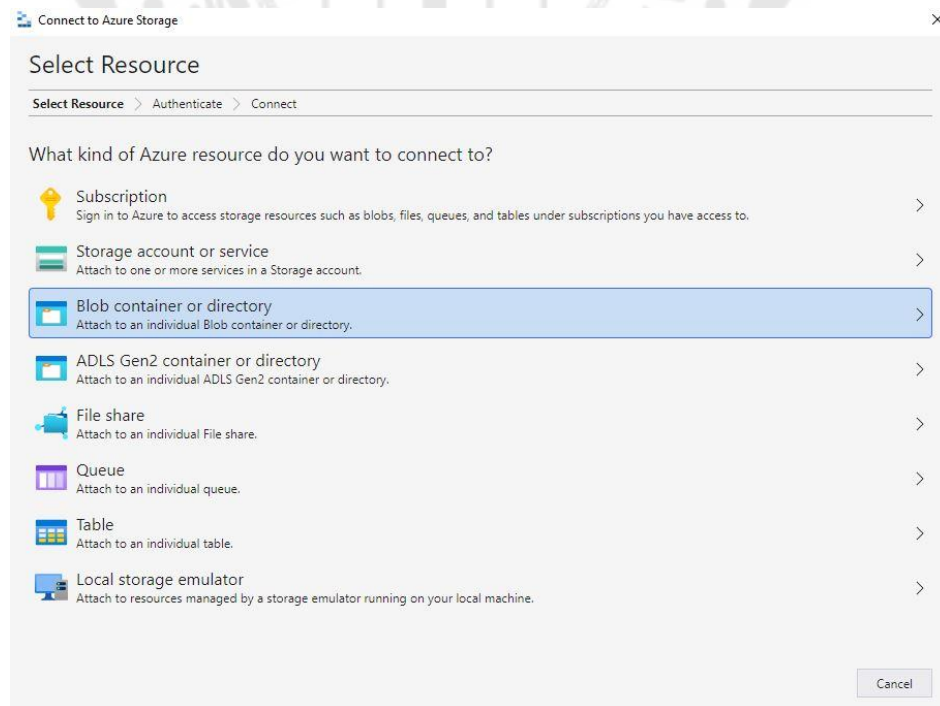
ภาพประกอบ 45 แสดง Azure signature URL

ที่มา : <https://www.equinor.com/energy/volve-data-sharing>

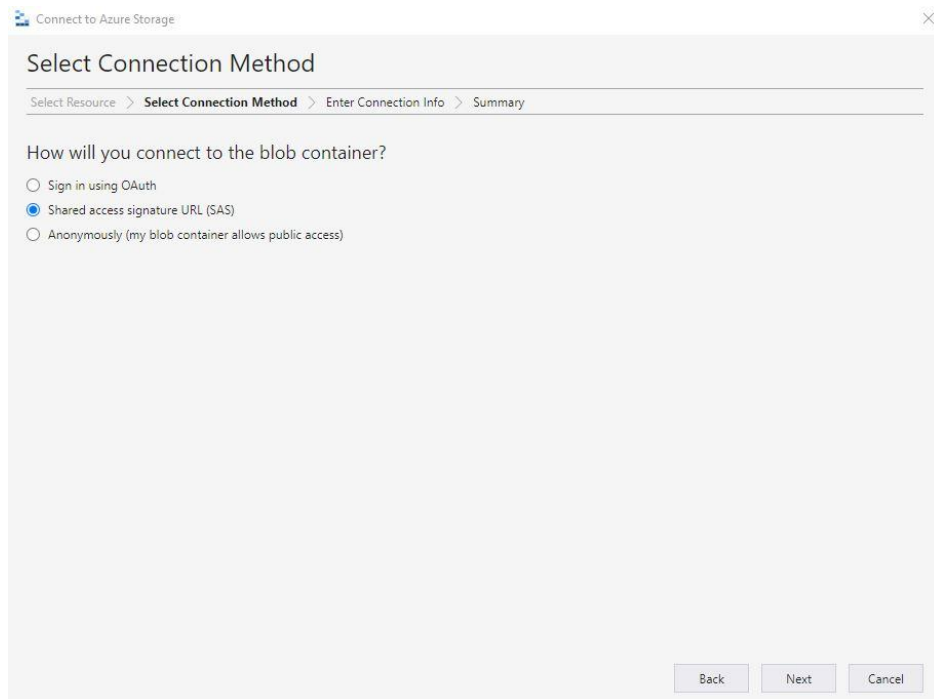
2. ใช้ Microsoft Azure Storage Explorer (โดยสามารถอ่านวิธีการดาวน์โหลด และการติดตั้งได้ที่ <https://azure.microsoft.com/en-us/products/storage/storage-explorer/>) เพื่อดาวน์โหลดชุดข้อมูล Volve Data Village โดยเมื่อเปิด Microsoft Azure Storage Explorer ขึ้นมาแล้ว ให้เลือกที่ Azure for Students คลิกขวาเลือก Storage Accounts เพื่อจะทำการ Connect to Azure Storage โดยการคลิกขวาแล้วเลือก Blob Container or directory เลือก Shared access signature URL (SAS) แล้วนำ Azure signature URL ที่ได้จากข้อ 1 มาใส่ใน Enter Connection Info หลังจากนั้นจะได้ชุดข้อมูลของ Volve Data Village ทั้งหมดเพื่อทำการดาวน์โหลด ดังภาพประกอบ 46 ถึง 49



ภาพประกอบ 46 แสดง Azure signature URL



ภาพประกอบ 47 แสดง Azure signature



Connect to Azure Storage

Select Connection Method

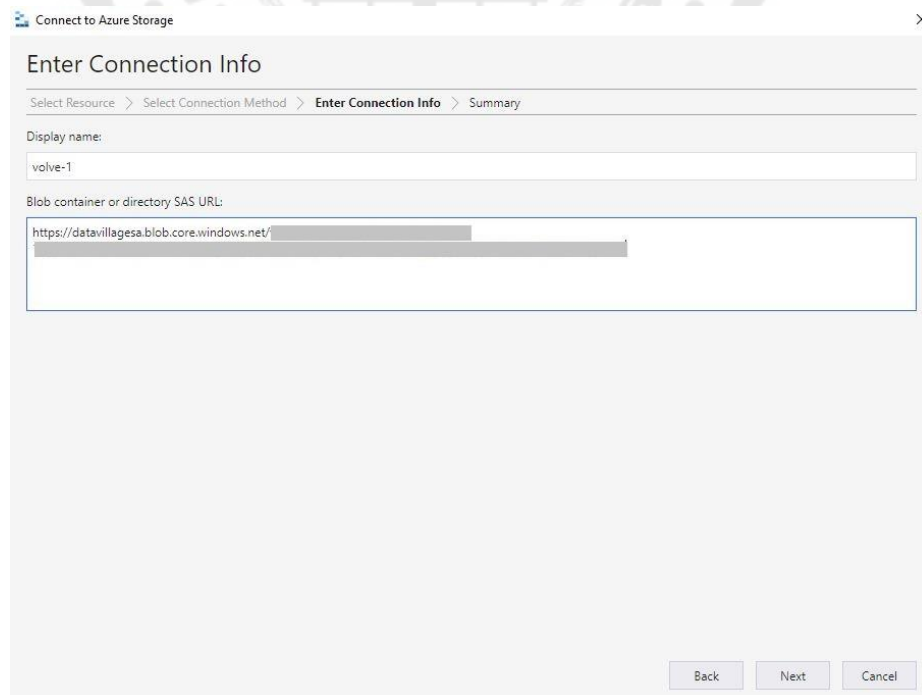
Select Resource > **Select Connection Method** > Enter Connection Info > Summary

How will you connect to the blob container?

- Sign in using OAuth
- Shared access signature URL (SAS)
- Anonymously (my blob container allows public access)

Back Next Cancel

ภาพประกอบ 48 แสดง Azure signature URL



Connect to Azure Storage

Enter Connection Info

Select Resource > Select Connection Method > **Enter Connection Info** > Summary

Display name:

volve-1

Blob container or directory SAS URL:

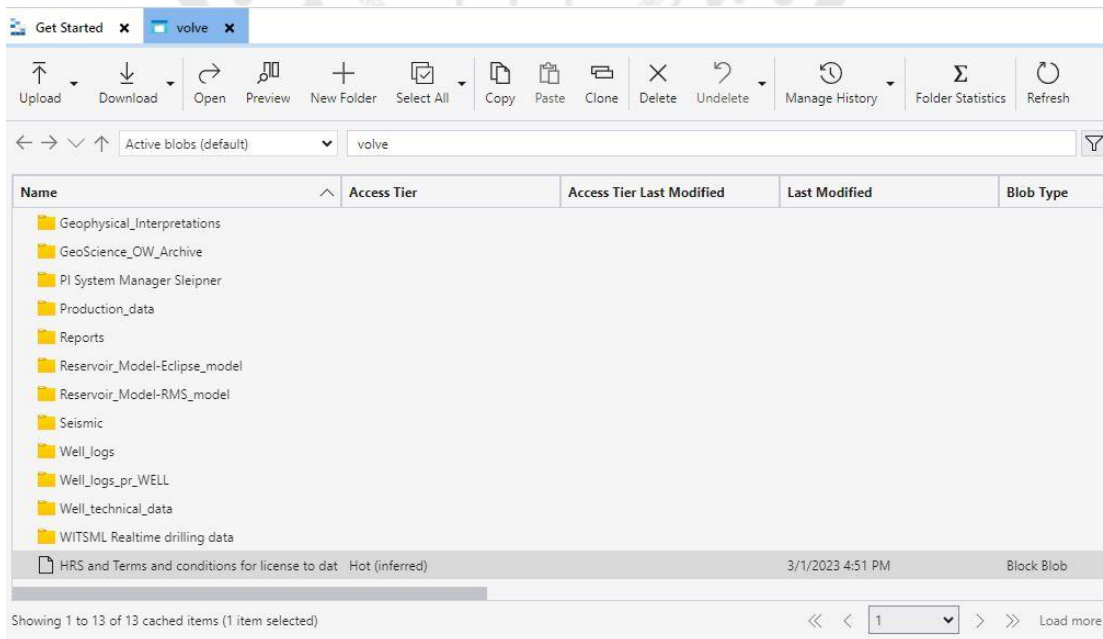
https://datavillagesa.blob.core.windows.net/

Back Next Cancel

ภาพประกอบ 49 แสดง Azure signature URL

ชุดข้อมูล Volve Data Village ภายในประกอบด้วย 12 folders ดังภาพประกอบ 50 ทั้งหมด 5 TB ประมาณ 40,000 files เกือบทุก folders จะมีไฟล์ readme โดย 12 folders ประกอบด้วย

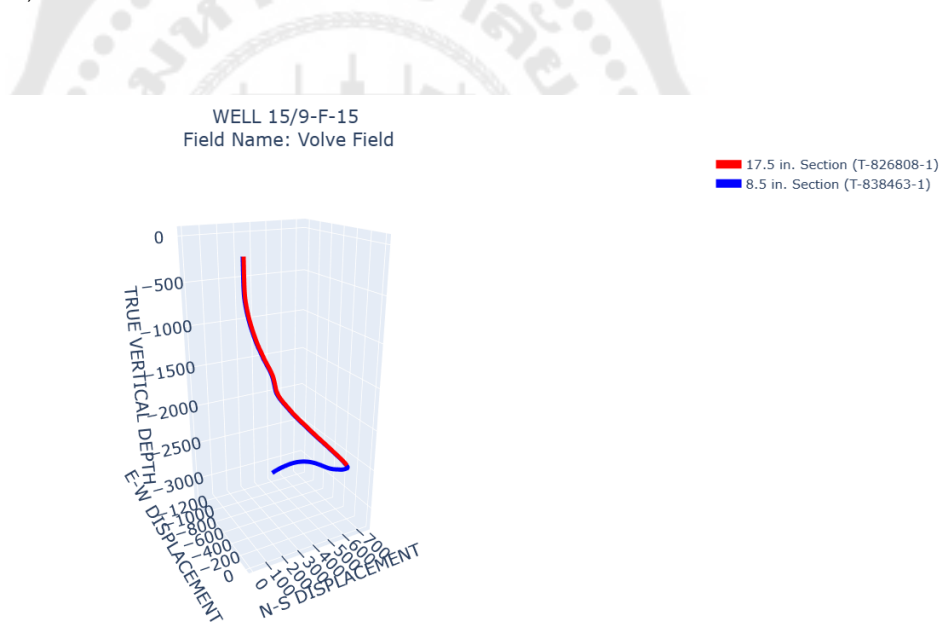
1. Geophysical Interpretations
2. GeoScience_OW_Archive
3. PI System Manager Sleipner
4. Production data
5. Reports
6. Reservoir_Model-Eclipse_model
7. Reservoir_Model-RMS_model
8. Seismic
9. Well_logs
10. Well_logs_pr_WELL
11. Well_technical_data
12. WITSMML Realtime drilling data (ใช้ชุดข้อมูลนี้เพื่อการทำกรวิจัย)



ภาพประกอบ 50 แสดงชุดข้อมูลทั้งหมดภายใน Volve Data Village

3.2.3 การทำความสะอาดและเตรียมข้อมูล (Data Cleaning and Preprocessing)

ข้อมูลที่ใช้ทำการวิจัยจะใช้ข้อมูลใน WITSML Realtime drilling data folder ทำการเลือกหลุม 15/9-F-15, StatoilHydro (Norway-StatoilHydro-15_\$47\$_9-F-15) ข้อมูลในการขุดเจาะแบ่งออกเป็น 2 ส่วน ได้แก่ 17.5" section และ 8.5" section ถูกเก็บใน Wellsite information transfer standard markup language (WITSML) นำเข้าและวิเคราะห์โดย Python ขั้นตอนการเข้าถึงข้อมูลและการนำเข้าข้อมูลด้วย Python สามารถศึกษาเพิ่มเติมได้ที่เว็บไซต์ DiscoverVolve [<https://discovervolve.com/2021/01/06/understanding-the-witsml-format-in-the-volve-oilfield-dataset/>] และ Github [<https://github.com/maribickpostanes/Interactive-Well-Trajectory-Plot>] เป็นหลุมเจาะแบบสำรวจ (Exploration well) โดยมีการเจาะเอียงทำมุม (Deviation well) ดังภาพประกอบที่ 51



ภาพประกอบ 51 แสดงทิศทางการเจาะของหลุมเจาะเอียงไปจากแนวตรง (Deviation Well)

ข้อมูลหลุม 15/9-F-15 เป็นข้อมูลที่เก็บแบบละเอียด เมื่อทำการสำรวจข้อมูลในเบื้องต้นจะพบข้อมูลจำนวนแถวทั้งหมด 112,714 แถว และมีจำนวนคอลัมน์ทั้งหมด 189 คอลัมน์ โดยข้อมูลดิบที่ได้มานั้นจะไม่ได้ใช้ทุกคอลัมน์ จึงทำการคัดเลือกคอลัมน์ที่น่าจะเกี่ยวข้องออกมา

ในเบื้องต้นก่อน ทำให้เหลือข้อมูลจำนวนแถวทั้งหมด 112,714 แถว และมีคอลัมน์ทั้งหมด 24 คอลัมน์ ดังภาพประกอบ 52

Measured Depth #	Hole depth (MD) #	Hole Depth (TVD) #	Total Vertical Depth #	Rate of Penetration #/h	Rate of Penetration (Sft avg) #/h	Weight on Bit #kgf	Weight on Bit #driller t	Average Rotary Speed #rpm	Average RPM #rpm	Average Surface Torque #kN.m	Average Torque #kN.m	Average Standpipe Pressure #kPa	Stand Pipe Pressure #kPa	Flow Pump #L/min	Mud Flow In #L/min	Mud Weight #g/cc3	Bit Drill Time #h	Bit Drilling Time #s	Bit Drilling Run #	Revolutions (cum) unitless	Bit Pressure Gradient #g/cc3	FPMD Pressure Gradient (active) #g/cc3	Tank volume (active) #m3	name	
0	1178.900	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
1	1178.902	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
2	1178.814	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
3	1178.905	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
4	1179.110	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log

ภาพประกอบ 52 แสดงการคัดเลือกคอลัมน์ในเบื้องต้น

จากนั้นได้ทำการตัดข้อมูลความลึก (Measure Depth) ที่ไม่ได้มีการเก็บข้อมูลไว้ โดยเริ่มตัดความลึกตั้งแต่เริ่มต้นจนถึง 1200 เมตร ออกจากข้อมูลทั้งหมด ดังภาพประกอบที่ 53 และทำการตัดข้อมูลที่เกิดความซ้ำซ้อนกันออก เนื่องจากใน 17.5"Section ได้ทำการเจาะความลึกถึง 2591.141 เมตร ทำให้ข้อมูลก่อน 8.5" section ซึ่งเป็นการลงก้านเจาะก่อนเจาะจริงเกิดการซ้ำซ้อนกันเกิดขึ้นในช่วงความลึก 2,464.156 เมตร ถึง 2591.105 เมตร จึงได้ทำการตัดข้อมูลที่ซ้ำซ้อนออกจากข้อมูลทั้งหมด และทำการลบคอลัมน์ที่ไม่เกี่ยวข้องอีก 8 คอลัมน์ ทำให้มีข้อมูลจำนวนแถวลดลงเหลือทั้งหมด 111,134 แถว และมีคอลัมน์ทั้งหมด 16 คอลัมน์

Measured Depth #	Hole depth (MD) #	Hole Depth (TVD) #	Total Vertical Depth #	Rate of Penetration #/h	Rate of Penetration (Sft avg) #/h	Weight on Bit #kgf	Weight on Bit #driller t	Average Rotary Speed #rpm	Average RPM #rpm	Average Surface Torque #kN.m	Average Torque #kN.m	Average Standpipe Pressure #kPa	Stand Pipe Pressure #kPa	Flow Pump #L/min	Mud Flow In #L/min	Mud Weight #g/cc3	Bit Drill Time #h	Bit Drilling Time #s	Bit Drilling Run #	Revolutions (cum) unitless	Bit Pressure Gradient #g/cc3	FPMD Pressure Gradient (active) #g/cc3	Tank volume (active) #m3	name	
0	1200.150	1200.1900	1181.350080	NaN	NaN	NaN	1.310882	NaN	0.0	NaN	0.013558	NaN	236.785995	NaN	NaN	0.0	NaN	137.56	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
1	1200.302	1200.3024	1181.505384	NaN	NaN	NaN	1.342033	NaN	0.0	NaN	0.013558	NaN	236.785995	NaN	NaN	0.0	NaN	137.57	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
2	1200.455	1200.4548	1181.651688	NaN	783.380872	783.380872	1.229235	NaN	0.0	NaN	0.013558	NaN	236.785995	NaN	NaN	0.0	NaN	137.57	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
3	1200.607	1200.6072	1181.797992	NaN	784.688912	784.688912	1.288202	NaN	0.0	NaN	0.013558	NaN	236.785995	NaN	NaN	0.0	NaN	137.57	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log
4	1200.760	1200.7596	1181.944296	NaN	786.914352	786.914352	1.295451	NaN	0.0	NaN	0.013558	NaN	236.785995	NaN	NaN	0.0	NaN	137.57	NaN	NaN	NaN	NaN	NaN	NaN	17.5m Section -MD Log

ภาพประกอบ 53 แสดงการตัดช่วงความลึกที่ไม่ได้มีการเก็บข้อมูลไว้ ออกจากข้อมูลทั้งหมด

หลังจากนั้นได้ทำการแยกข้อมูลตัวหนังสือกับตัวเลขออกจากกันในคอลัมน์ name และแทนที่ด้วยค่าตัวเลขเพียงอย่างเดียว ดังภาพประกอบ 54

Measured Depth m	Rate of Penetration m/h	Rate of Penetration (5ft avg) m/h	Weight on Bit kkgf	Average Rotary Speed rpm	Average Surface Torque kN.m	Average Standpipe Pressure kPa	Flow Pumps L/min	Mud Flow In L/min	Mud Weight Out g/cm3	Bit Drill Time h	Bit Drilling Time s	Bit Drilling Run m	Bit Revolutions (cum) unitless	FPWD Pore Pressure Gradient g/cm3	name
0	1200.150	NaN	NaN	1.310882	0.0	0.013558	236.765965	NaN	0.0	NaN	137.56	NaN	NaN	NaN	17.5
1	1200.302	NaN	NaN	1.342633	0.0	0.013558	236.765965	NaN	0.0	NaN	137.57	NaN	NaN	NaN	17.5
2	1200.455	793.360872	793.360872	1.229235	0.0	0.013558	236.765965	NaN	0.0	NaN	137.57	NaN	NaN	NaN	17.5
3	1200.607	784.689312	784.689312	1.288202	0.0	0.013558	236.765965	NaN	0.0	NaN	137.57	NaN	NaN	NaN	17.5
4	1200.760	786.914352	786.914352	1.256451	0.0	0.013558	236.765965	NaN	0.0	NaN	137.57	NaN	NaN	NaN	17.5

ภาพประกอบ 54 แสดงการแยกข้อมูลตัวหนังสือและตัวเลขออกจากกันและแทนที่ด้วยค่าตัวเลขเพียงอย่างเดียว

ทำการเติมค่าช่องว่าง (NaN) ด้วยวิธีเติมค่าข้อมูลด้วยวิธี Forward Fill และ Backward Fill จากนั้นจะพบว่ายังมีค่าข้อมูลติดลบอยู่ใน 4 คอลัมน์ ได้แก่ 1. Weight on Bit 2. Surface Torque 3. Flow Pump ทำการแทนค่าติดลบด้วยค่า 0 และ 4. Mud Weight ทำการแทนค่าติดลบด้วยค่า 11.933929 และ 11.850475 (นำค่ามาจากรายงานการเจาะหลุม) ดังภาพประกอบ 55 ถึง 56

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111134 entries, 0 to 111133
Data columns (total 16 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Measured Depth m                         111134 non-null float64
1   Rate of Penetration m/h                  111134 non-null float64
2   Rate of Penetration (5ft avg) m/h       111134 non-null float64
3   Weight on Bit kkgf                       111134 non-null float64
4   Average Rotary Speed rpm                 111134 non-null float64
5   Average Surface Torque kN.m             111134 non-null float64
6   Average Standpipe Pressure kPa          111134 non-null float64
7   Flow Pumps L/min                        111134 non-null float64
8   Mud Flow In L/min                       111134 non-null float64
9   Mud Weight Out g/cm3                    111134 non-null float64
10  Bit Drill Time h                         111134 non-null float64
11  Bit Drilling Time s                      111134 non-null float64
12  Bit Drilling Run m                       111134 non-null float64
13  Bit Revolutions (cum) unitless           111134 non-null float64
14  FPWD Pore Pressure Gradient g/cm3       111134 non-null float64
15  name                                     111134 non-null float64
dtypes: float64(16)
memory usage: 13.6 MB
```

ภาพประกอบ 55 แสดงการเติมค่าว่าง (NaN) ทั้งหมดในข้อมูล

	count	mean	std	min	25%	50%	75%	max
Measured Depth m	111134.0	2.513851e+03	7.734257e+02	1200.150000	1836.868000	2463.428500	3142.949500	4.065346e+03
Rate of Penetration m/h	111134.0	2.524718e+01	4.298708e+01	0.786384	15.266518	19.671792	24.885091	8.103961e+02
Rate of Penetration (5ft avg) m/h	111134.0	2.480611e+01	4.194384e+01	1.847088	14.938248	19.674840	24.871680	8.076286e+02
Weight on Bit kkgf	111134.0	5.744417e+00	6.348784e+00	-2.685267	1.923232	4.578108	8.334760	8.698994e+01
Average Rotary Speed rpm	111134.0	1.305520e+02	1.987593e+01	0.000000	129.390000	139.640000	139.736000	1.430500e+02
Average Surface Torque kNm	111134.0	1.784067e+01	7.677005e+00	-333.748146	12.080338	17.042632	23.599367	3.648913e+01
Average Standpipe Pressure kPa	111134.0	6.395786e+27	6.982077e+29	205.463767	15145.713350	20326.158190	25656.839790	1.300000e+32
Flow Pumps L/min	111134.0	3.372227e+03	1.152180e+03	-535.469971	2092.020021	4028.979982	4451.779784	4.490720e+03
Mud Flow In L/min	111134.0	3.349487e+03	1.163824e+03	0.000000	2114.455435	4010.000494	4453.120823	4.573043e+03
Mud Weight Out g/cm3	111134.0	-9.242238e+00	2.232734e+02	-2426.593299	11.182843	11.767021	12.434650	1.243465e+01
Bit Drill Time h	111134.0	2.017458e+02	2.793582e+01	137.560000	179.362500	215.555000	225.667000	2.256670e+02
Bit Drilling Time s	111134.0	1.080511e+05	7.629634e+04	834.429993	43239.218750	94777.632810	154170.015600	2.878508e+05
Bit Drilling Run m	111134.0	5.154156e+02	3.466006e+02	0.950000	221.399994	480.489990	760.429993	1.269420e+03
Bit Revolutions (cum) unitless	111134.0	4.840380e+05	3.501796e+05	0.000000	193658.000000	420529.000000	701478.000000	1.303268e+06
FPWD Pore Pressure Gradient g/cm3	111134.0	9.586114e-01	4.440912e-16	0.958611	0.958611	0.958611	0.958611	9.586114e-01
name	111134.0	1.356729e+01	4.464120e+00	8.500000	8.500000	17.500000	17.500000	1.750000e+01

ภาพประกอบ 56 แสดงค่าติดลบที่อยู่ในข้อมูล

จากนั้นทำทศนิยมในค่าข้อมูลความลึกให้เป็น 0 และทำการลบข้อมูลค่าความลึกที่ซ้ำซ้อนและเก็บไว้เพียงค่าแรกให้เป็นตัวแทนข้อมูลในค่าความลึกนั้นๆ และทำการเพิ่มข้อมูลคอลัมน์ Total Flow Area (TFA) โดย 17.5" Section เพิ่มค่าข้อมูลเป็น 1.633 และ 8.5" Section เพิ่มค่าข้อมูลเป็น 0.721 และ 0.509 ตามลำดับ โดยดูจาก Bit No. ในการแทนค่า 8.5" Section ทำให้มีข้อมูลจำนวนแถวลดลงเหลือทั้งหมด 2,701 แถว และมีคอลัมน์ทั้งหมด 17 คอลัมน์ ดังภาพประกอบที่ 57 ถึง 58

	Measured Depth m	Rate of Penetration m/h	Rate of Penetration (5ft avg) m/h	Weight on Bit kkgf	Average Rotary Speed rpm	Average Surface Torque kN.m	Average Standpipe Pressure kPa	Flow Pumps L/min	Mud Flow In L/min	Mud weight Out g/cm3	Bit Drill Time h	Bit Drilling Time s	Bit Drilling Run m	Bit Revolutions (cum) unitless	FPWD Pore Pressure Gradient g/cm3	name
0	1200.0	793.360072	793.360872	1.310882	0.0	0.013558	236.765965	3825.500001	0.0	11.933929	137.56	1845.819946	0.95	3035.0	0.958611	17.5
1	1201.0	784.689312	784.689312	1.288202	0.0	0.013558	236.765965	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5
2	1202.0	784.646640	784.646640	1.324490	0.0	0.013558	213.875371	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5
3	1203.0	764.843784	779.711928	1.465103	0.0	0.013558	213.875371	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5
4	1204.0	4.212336	20.464272	2.013950	0.0	0.188815	224.493298	3825.500001	0.0	11.933929	137.64	1845.819946	0.95	3035.0	0.958611	17.5
5	1205.0	231.757728	19.635216	1.896016	0.0	0.203373	222.218026	3825.500001	0.0	11.933929	137.65	1845.819946	0.95	3035.0	0.958611	17.5
6	1206.0	398.074896	22.716744	2.054773	0.0	0.325396	241.730191	3825.500001	0.0	11.933929	137.71	1845.819946	0.95	3035.0	0.958611	17.5
7	1207.0	583.858472	23.518368	2.304249	0.0	0.325396	241.730191	3825.500001	0.0	11.933929	137.71	1845.819946	0.95	3035.0	0.958611	17.5
8	1208.0	626.251224	608.469192	2.263426	0.0	0.338954	255.657800	3825.500001	0.0	11.933929	137.71	1845.819946	0.95	3035.0	0.958611	17.5
9	1209.0	627.171720	625.336824	2.240746	0.0	0.338954	255.657800	3825.500001	0.0	11.933929	137.72	1845.819946	0.95	3035.0	0.958611	17.5

ภาพประกอบ 57 แสดงค่าติดลบที่อยู่ในข้อมูล

Measured Depth m	Rate of Penetration m/h	Rate of Penetration (5ft avg) m/h	Weight on Bit kkgf	Average Rotary Speed rpm	Average Surface Torque kN.m	Average Standpipe Pressure kPa	Flow Pumps L/min	Mud Flow In L/min	Mud weight Out g/cm3	Bit Drill Time h	Bit Drilling Time s	Bit Drilling Run m	Revolutions (cum) unitless	Bit Pressure Gradient g/cm3	name	Total Flow Area (TFA) in2	
0	1200.0	793.360872	793.360872	1.310882	0.0	0.013558	236.765965	3825.500001	0.0	11.933929	137.56	1845.819946	0.95	3035.0	0.958611	17.5	1.633
1	1201.0	784.689312	784.689312	1.288202	0.0	0.013558	236.765965	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5	1.633
2	1202.0	784.646640	784.646640	1.324490	0.0	0.013558	213.875371	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5	1.633
3	1203.0	764.843784	779.711928	1.465103	0.0	0.013558	213.875371	3825.500001	0.0	11.933929	137.57	1845.819946	0.95	3035.0	0.958611	17.5	1.633
4	1204.0	4.212336	20.464272	2.013950	0.0	0.189815	224.493298	3825.500001	0.0	11.933929	137.64	1845.819946	0.95	3035.0	0.958611	17.5	1.633

ภาพประกอบ 58 แสดงการเพิ่มคอลัมน์ Total Flow Area

3.2.4 การเลือกค่าพารามิเตอร์ที่เกี่ยวข้อง (Selection Parameters)

ทำการเลือกข้อมูลคอลัมน์ที่เป็นข้อมูลปัจจัยที่สำคัญและเกี่ยวข้อง จำนวน 10 คอลัมน์ ดังภาพประกอบ 59 ประกอบด้วย

1. Rate of Penetration (ROP)
2. Measure Depth (MD)
3. Weight on Bit (WOB)
4. Rotary Speed (Revolution per minute, RPM)
5. Surface Torque (T)
6. Standpipe Pressure (SPP)
7. Mud Flow In (Q)
8. Mud Weight (MW)
9. Bit Diameter (D_{BIT}) จาก name เดิม
10. Total Flow Area (TFA)

	Rate of Penetration	Measure Depth	Weight on Bit	Rotary Speed	Surface Torque	Standpipe Pressure	Mud Flow In	Mud Weight	Bit Diameter	Total Flow Area
0	793.360872	1200.0	1.310882	0.0	0.013558	236.765965	0.0	11.933929	17.5	1.633
1	784.689312	1201.0	1.288202	0.0	0.013558	236.765965	0.0	11.933929	17.5	1.633
2	784.646640	1202.0	1.324490	0.0	0.013558	213.875371	0.0	11.933929	17.5	1.633
3	764.843784	1203.0	1.465103	0.0	0.013558	213.875371	0.0	11.933929	17.5	1.633
4	4.212336	1204.0	2.013950	0.0	0.189815	224.493298	0.0	11.933929	17.5	1.633

ภาพประกอบ 59 แสดงรายละเอียดข้อมูลคอลัมน์ที่สำคัญและเกี่ยวข้อง

3.3 การสำรวจข้อมูลคุณลักษณะเบื้องต้น (Exploratory Data Analysis)

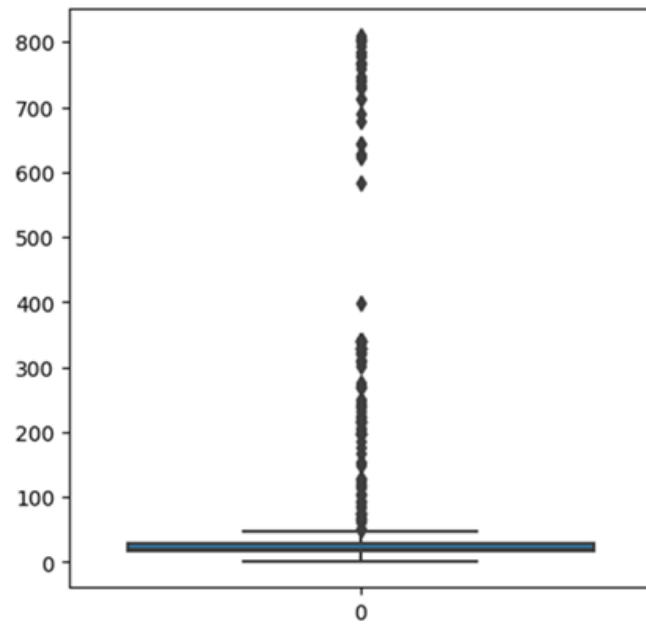
เมื่อทำการสำรวจข้อมูลเบื้องต้นจากข้อมูลจะพบจำนวนแถว 2,701 แถว และจำนวนคอลัมน์ 10 คอลัมน์ ดังภาพประกอบ 60

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2701 entries, 0 to 2700
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Rate of Penetration    2701 non-null   float64
1   Measure Depth          2701 non-null   float64
2   Weight on Bit          2701 non-null   float64
3   Rotary Speed           2701 non-null   float64
4   Surface Torque         2701 non-null   float64
5   Standpipe Pressure     2701 non-null   float64
6   Mud Flow In            2701 non-null   float64
7   Mud Weight             2701 non-null   float64
8   Bit Diameter           2701 non-null   float64
9   Total Flow Area        2701 non-null   float64
dtypes: float64(10)
memory usage: 211.1 KB
```

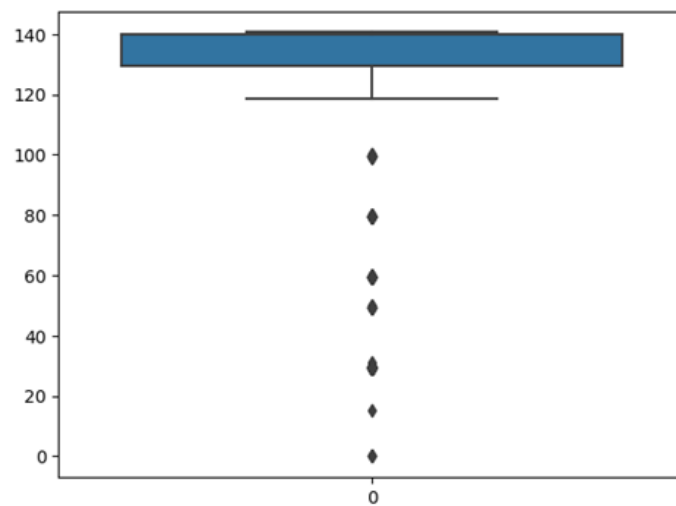
ภาพประกอบ 60 แสดงรายละเอียดข้อมูลเบื้องต้นทั้งหมด

3.3.1 ข้อมูลที่ผิดปกติ (Outliers)

ทำการตรวจสอบค่าข้อมูลความผิดปกติของข้อมูลการเจาะจะเห็นว่าค่าข้อมูลความผิดปกติอยู่ที่ค่ามากกว่า 400 และค่าข้อมูลความผิดปกติของ Rotary Speed ที่มีค่าเท่ากับ 0 ดังภาพประกอบที่ 61 ถึง 62 จึงทำการตัดค่าข้อมูลผิดปกติเหล่านี้ออกไป จำนวนข้อมูลแถวที่เหลือมีทั้งหมด 2,664 แถว คอลัมน์ทั้งหมด 10 คอลัมน์ และค่าข้อมูลรวมทางสถิติหลังจากตัดข้อมูลที่ผิดปกติออก ดังตารางที่ 2 และ 3



ภาพประกอบ 61 แสดงค่าความผิดปกติของค่าอัตราการเจาะ



ภาพประกอบ 62 แสดงค่าความผิดปกติของค่า Rotary Speed

ตาราง 2 สรุปข้อมูลรวมทางสถิติของข้อมูล Rate of Penetration, Measure Depth, Weight on Bit, Rotary Speed, Surface Torque

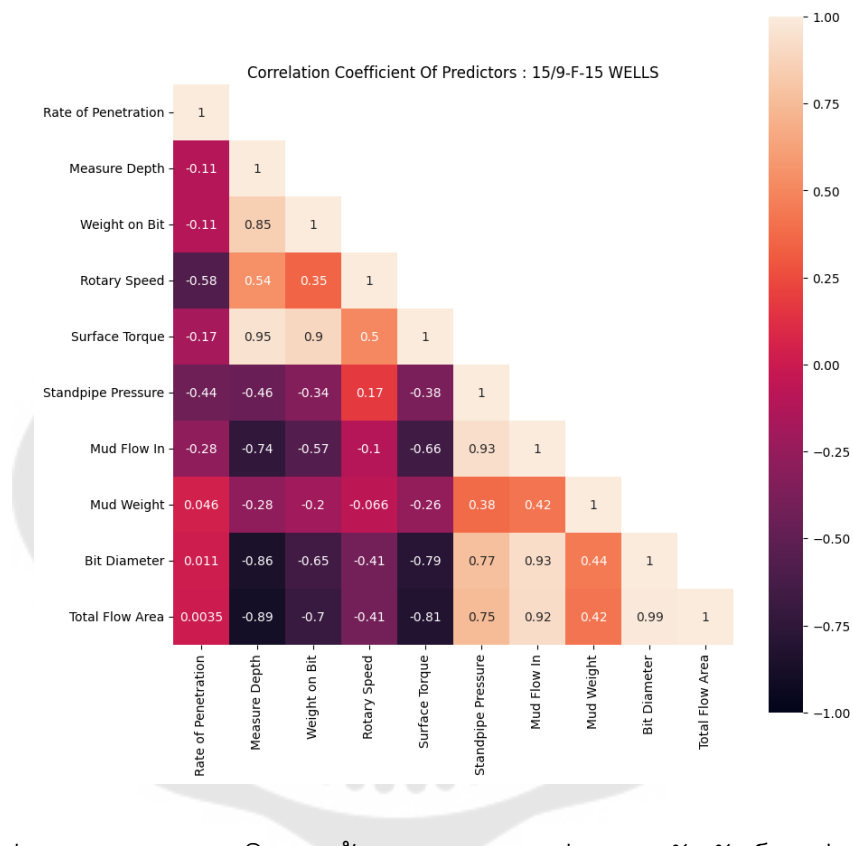
Statistical Parameters	ROP (m/hr)	MD (m)	WOB (kkgf)	RPM (rpm)	T (kN.m)
Maximum	340.52	4065.00	19.86	140.57	35.19
Minimum	1.62	1237.00	0.00	15.33	1.10
Range	338.90	2828.00	19.86	125.24	34.09
Mean	29.16	2602.50	6.05	129.87	18.77
Median	19.89	2568.50	5.06	139.74	17.86
Standard Deviation	41.71	817.33	4.52	23.56	7.47

ตาราง 3 สรุปข้อมูลรวมทางสถิติของข้อมูล Standpipe Pressure, Mud Flow In, Mud Weight out, Bit Daimeter, Total Flow Area

Statistical Parameters	SPP (kPa)	Q (L/min)	MW _{out} (g/cm ³)	D _{BIT} (Inch)	TFA (Inch ²)
Maximum	27718.94	4475.27	12.43	17.50	1.63
Minimum	4330.53	1083.31	4.84	8.50	0.51
Range	23388.41	3391.96	7.59	9.00	1.12
Mean	19358.00	3181.45	11.37	13.08	1.15
Median	16364.20	2121.38	11.43	17.50	1.63
Standard Deviation	5845.70	1167.11	0.91	4.50	0.49

3.3.2 การหาค่าความสัมพันธ์ระหว่างตัวแปร (Parameters Correlation)

ทำการหาค่าความสัมพันธ์ระหว่างค่าข้อมูลของตัวแปร โดยแผนภูมิความร้อน (Heatmap) ค่าตัวเลขในแผนภูมิความร้อนยังมีค่าเข้าใกล้ 1 แสดงถึง ค่าความสัมพันธ์ระหว่างตัวแปรที่มีความสัมพันธ์กันมาก และ ถ้าค่าตัวเลขในแผนภูมิความร้อนยังมีค่าเข้าใกล้ -1 แสดงถึง ค่าความสัมพันธ์ระหว่างตัวแปรที่มีความสัมพันธ์กันในทางตรงกันข้าม ดังภาพประกอบ 63



ภาพประกอบ 63 แผนภูมิความร้อนแสดงการหาค่าความสัมพันธ์ระหว่างตัวแปร

3.3.3 การอธิบายค่าตัวแปร (Explain Feature)

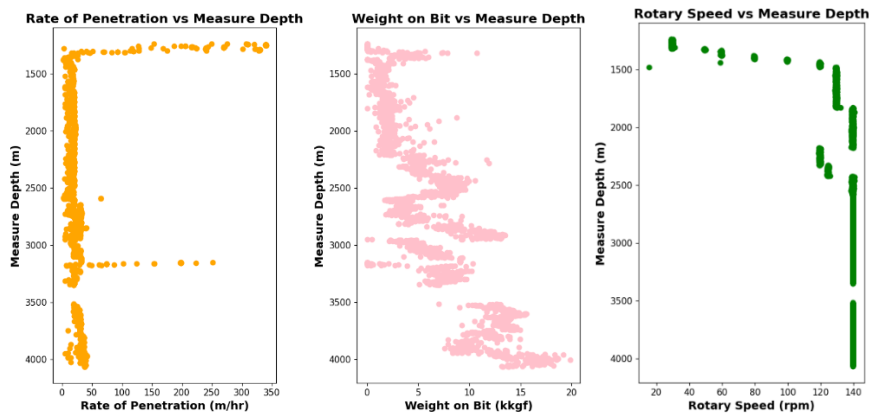
ที่ค่าความลึกระหว่าง 2,592 ถึง 2,604 เมตร พบว่าค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) มีค่าสูงผิดปกติเนื่องจากการเปลี่ยนแปลงของชั้นหิน โดยทำการลบและแก้ไขด้วยวิธีเชิงเส้น (Linear Interpolation) จากภาพประกอบที่ 64 (ก) เป็นกราฟที่พล็อตระหว่างค่าความลึกที่วัดได้ (Measure Depth, MD) กับ ค่าอัตราการเจาะ (Rate of Penetration, ROP) ในช่วงเริ่มต้นค่าอัตราการเจาะจะมีค่าสูงที่ค่าความลึกระหว่าง 1,200 ถึง 1,306 เมตร หลังจากนั้น

ค่าอัตราการเจาะจะค่อยๆลดลงและเริ่มคงที่ที่ค่าความลึก 1,320 เมตร ในส่วน 17.5"section ค่าอัตราการเจาะอยู่ระหว่าง 5 ถึง 25 ขณะนี้ 8.5"section จะมีค่าอยู่ระหว่าง 20 ถึง 40 โดยบางช่วงจะมีการการเพิ่มอย่างรวดเร็ว เช่น ที่ค่าความลึก 3,151 ถึง 3,177 เมตร เนื่องมาจากการเปลี่ยนแปลงของชั้นหิน ค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) และ ความเร็วในการหมุนก้านเจาะ (Rotary Speed, RPM) เป็นปัจจัยที่ส่งผลต่ออัตราการเจาะเป็นอย่างมาก ในภาพประกอบที่ 64 (ข) จะสังเกตได้ว่าค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) ใน 17.5"section จะมีค่าอยู่ระหว่าง 0 ถึง 5 ซึ่งเป็นช่วงต้นของการเจาะ จากนั้นเพิ่มเป็น 5 ถึง 11 อยู่ที่ค่าความลึก 2,213 ถึง 2,591 เมตร สำหรับ 8.5"section ค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) จะเพิ่มสูงจนถึง 13 จากนั้นจะลดลงในช่วงค่าความลึก 2,591 ถึง 3,517 เมตร โดยค่าเฉลี่ยของค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) ช่วงนี้จะอยู่ที่ 6 หลังจากนั้นจะลดค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) และจะเพิ่มค่าน้ำหนักที่กดลงบนหัวเจาะให้อยู่ระหว่าง 7 ถึง 19 ค่าอัตราการรอบของการหมุนของก้านเจาะ (Rotary Speed, RPM) จากภาพประกอบที่ 64 (ค) ใน 17.5"section ค่าอัตราการรอบของการหมุนของก้านเจาะค่อยๆเพิ่มขึ้นจนถึงที่ค่าความลึก 1,480 เมตร หลังจากนั้นค่าอัตราการรอบของการหมุนของก้านเจาะจะคงที่ที่ 130 และ 8.5"section ค่าอัตราการรอบของการหมุนของก้านเจาะประมาณ 140

จากภาพประกอบ 64 (ง) ใน 17.5"section ค่าน้ำโคลนที่ออกจากหลุมจะถูกรักษาให้อยู่ที่ประมาณ 11.8-12.0 จะมีบางช่วงที่ค่าน้ำโคลนที่ออกมามีค่าต่ำลง แต่ก็พยายามรักษาให้กลับมาอยู่ในระดับที่ 11.8-12.0 ขณะนี้ 8.5 "section ช่วงแรกค่าน้ำโคลนจะอยู่ที่ 10.68-10.75 จนถึงค่าความลึกที่ 2930 เมตร น้ำหนักน้ำโคลนที่ออกมาจะถูกรักษาระดับให้เพิ่มขึ้นอยู่ระดับที่ 11.1-11.5 จากภาพประกอบ 64 (จ) ค่า Standpipe Pressure จะแบ่งชัดเจนระหว่าง 2 section ใน 17.5"section จะมีค่าสูงกว่า 8.5"section โดย 17.5"section ค่าจะอยู่ช่วง 20,000 และค่อยๆเพิ่มขึ้นจนถึง 27,500 ขณะนี้ 8.5"section ค่าจะอยู่ที่ช่วง 14,000-16,000 โดยเป็นการค่อยๆเพิ่มเช่นเดียวกับ section แรก จากภาพประกอบ 64 (ฉ) ค่า Mud Flow In จะแบ่งชัดเจนระหว่าง 2 section คือ 17.5"section ค่าจะรักษาระดับให้อยู่ที่ 4,400 และใน 8.5 "section ค่าจะรักษาระดับให้อยู่ที่ 2,100 จนถึงค่าความลึกที่ 3,175 เมตร หลังจากนั้น จะลดลงมาเหลือ 2,000

จากภาพประกอบ 64 (ญ) ค่า TFA จะเปลี่ยนตามการเปลี่ยนหัวเจาะ โดย 17.5"section จะมีค่า 1.633 และ 8.5"section จะมีค่า 0.721 และ 0.509 ตามลำดับ จากภาพประกอบ 64 (ฐ) ค่า Surface Torque จะมีค่าเพิ่มขึ้นเรื่อยๆตั้งแต่ 17.5"section จนถึง

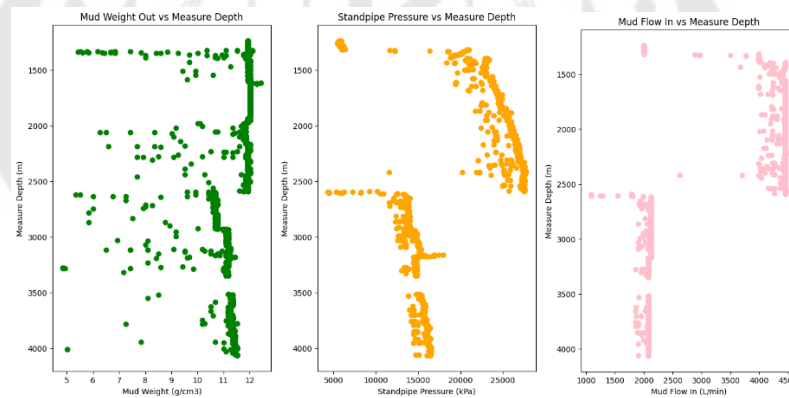
8.5" section โดยช่วงการเจาะ section แรกจะอยู่ 4-20 และช่วงการเจาะ section หลังจะอยู่ 15-35 จากภาพประกอบ 64 (ถ) ค่า Bit Diameter จะแบ่งได้เป็น 17.5 และ 8.5 ตามลำดับ



(ก)

(ข)

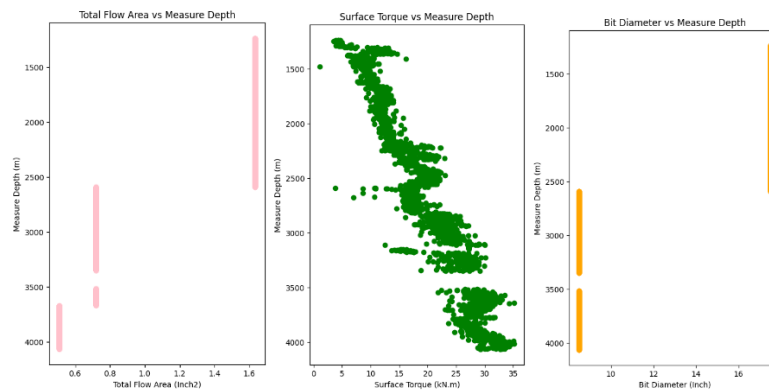
(ค)



(ง)

(จ)

(ฉ)



(ฐ)

(ฑ)

(ถ)

ภาพประกอบ 64 การอธิบายเปรียบเทียบปัจจัยกับช่วงค่าความลึกต่างๆที่วัดได้ (Measure Depth, MD) (ก) ค่าอัตราการเจาะ (Rate of Penetration, ROP) (ข) ค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) (ค) ค่าอัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM) (ง) ค่าน้ำหนักน้ำโคลนที่ออกมา (Mud Weight out, MW_{out}) (จ) ค่าความดันที่ก้านเจาะ (Standpipe Pressure, SPP) (ฉ) การไหลเข้าของน้ำโคลน (Mud Flow In, Q) (ญ) พื้นที่หน้าตัดรวมที่ทำให้ของไหลผ่านเข้าไปในหัวเจาะ (Total Flow Area, TFA) (ฐ) แรงบิดตัวด้านบน (Surface Torque, T) (ถ) ขนาดของหัวเจาะ (Bit Diameter, D_{bit})

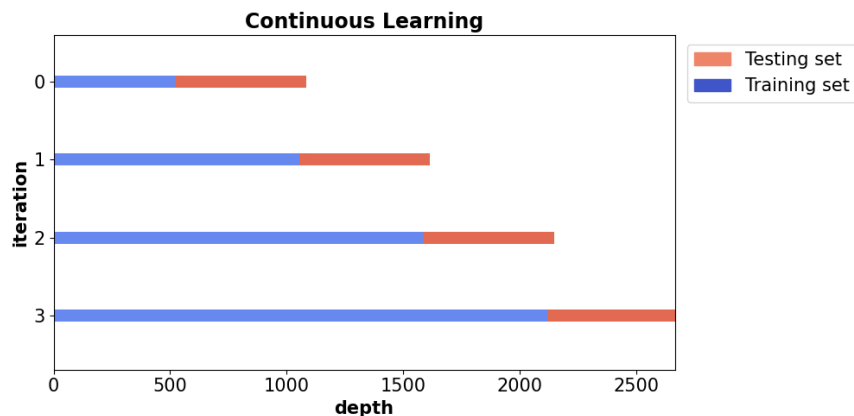
3.4 การตั้งค่าแบบจำลองและการฝึกฝน (Model setup and Training)

3.4.1 การแบ่งชุดข้อมูลการฝึกฝนและการทดสอบ (Train-Test Split)

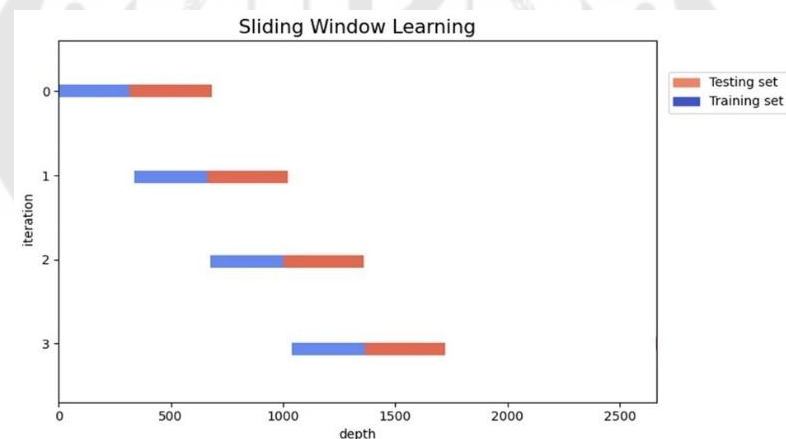
จากข้อมูลจำนวนแถวทั้งหมด 2,664 แถว และมีทั้งหมด 10 คอลัมน์ ทำการลบ คอลัมน์ Rate of Penetration ออก ให้เป็นค่าเป้าหมาย (Target) เพื่อใช้ในการทำนายค่าตัวแปร y และคอลัมน์ที่เหลืออีก 9 คอลัมน์เป็นค่าตัวแปร X โดยค่า X ใช้เป็นตัวแปรต้นในการทำนายค่าตัวแปรตาม y จากนั้นทำการแบ่งชุดข้อมูลเพื่อใช้ในการฝึกสอน (Training) และชุดข้อมูลเพื่อใช้ในการทดสอบ (Testing) ข้อมูลที่ใช้แบบตามลำดับข้อมูล (Sequential Data) หรือ ความต่อเนื่องของค่าความลึก (Depth-Dependent) ที่ใช้ในแบบจำลองของการดำเนินงานการขุดเจาะ ดังนั้นแบบจำลองที่ใช้ในการฝึกฝนและการทดสอบจะให้การแบ่งข้อมูลตามลำดับ (Sequential Split) โดย 2 เทคนิค ที่ใช้ คือ การแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or Forward-Chaining Split) และ การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) ทั้ง 2 เทคนิค แบ่งข้อมูลตามค่าความลึก ที่เรียกว่า การเพิ่มขนาด (Increments) ขึ้นของข้อมูล การเพิ่มขนาดขึ้นของข้อมูลเป็นตัวแทนของช่วงค่าความลึกที่ข้อมูลถูกแบ่งในแต่ละการวนซ้ำ (Iteration) ตัวอย่างของการเพิ่มขึ้น (Increments) ที่ใช้ในการวิจัย ได้แก่ 5, 10, 15, 20, 50, 100, 200 และ 500 เมตร เพื่อเลือกการทดลองที่เป็นตัวแทนการประเมินประสิทธิภาพแบบจำลองการเรียนรู้ที่ส่งผลต่อความแม่นยำ ทั้งตัวแปร X และ y โดยกำหนด Random Seed เท่ากับ 1234 ของ Numpy และ Random State ใน scikit-learn เพื่อให้แน่ใจว่าผลลัพธ์มีความสม่ำเสมอและสามารถทำซ้ำได้ [<https://sklearn-genetic-opt.readthedocs.io/en/stable/tutorials/reproducibility.html>]

การแบ่งข้อมูลแบบต่อเนื่องและการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน เริ่มแรกทั้ง 2 เทคนิคมีข้อมูลกรฝึกฝนและการทดสอบเท่ากัน แต่ในการวนซ้ำครั้งที่ 2 การเรียนรู้แบบต่อเนื่องจะรวมข้อมูลใหม่เข้ากับชุดข้อมูลฝึกฝนที่มีอยู่เดิม เช่น increments เท่ากับ 500 เมตร iteration ที่ 0 ซึ่งเป็นการฝึกฝนรอบที่ 1 มีข้อมูลการฝึกฝนที่ 500 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 0 ถึง 500 เมตร และข้อมูลทดสอบอีก 500 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 500 ถึง 1000 เมตร และ iteration ที่ 1 ในการวนซ้ำรอบที่ 2 การฝึกฝนมีข้อมูลที่ 1000 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 0 ถึง 1000 เมตร ข้อมูลการทดสอบอีก 500 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 1000 ถึง 1500 เมตร ดังภาพประกอบที่ 65 ทำการวนซ้ำแบบนี้ไปเรื่อยๆจนครบค่าความลึกของ increments นั้นๆ และทำการฝึกฝนการทดสอบจนครบแต่ละ increments ที่กำหนดไว้

ในขณะที่การแบ่งข้อมูลแบบหน้าต่างบานเลื่อนจะเลื่อนหน้าต่างของข้อมูลฝึกและข้อมูลทดสอบไปข้างหน้า โดยยังคงขนาด (Size) ของชุดข้อมูลฝึกและชุดข้อมูลทดสอบเท่ากับการเพิ่ม Increments ขึ้นของข้อมูล เช่น increments เท่ากับ 500 เมตร iteration ที่ 0 เป็นการฝึกฝนรอบที่ 1 มีข้อมูลการฝึกฝนที่ 500 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 0 ถึง 500 เมตร ข้อมูลการทดสอบอีก 500 เมตร อยู่ในช่วงค่าความลึกที่ระยะ 500 ถึง 1000 เมตร และ iteration ที่ 1 ในการวนซ้ำรอบที่ 2 นำข้อมูลการทดสอบเดิมจากรอบที่ 1 มาทำการฝึกฝนในรอบที่ 2 ซึ่งมีข้อมูลที่ 500 เมตร แต่อยู่ในช่วงค่าความลึกที่ระยะ 500 ถึง 1000 เมตร และข้อมูลทดสอบอีก 500 เมตร ในระยะความลึกที่ 1000 ถึง 1500 เมตร ดังภาพประกอบที่ 66 ทำการเลื่อนข้อมูลฝึกฝนข้อมูลทดสอบแบบนี้ไปเรื่อยๆจนครบค่าความลึกของ increments นั้นๆ และทำการฝึกฝนการทดสอบจนครบแต่ละ increments ที่กำหนดไว้ เช่นเดียวกันกับการเรียนรู้แบบต่อเนื่อง



ภาพประกอบ 65 การแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or forward-chaining split)



ภาพประกอบ 66 การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split)

3.4.2 การตั้งค่าแบบจำลอง (Model setup)

งานวิจัยนี้มีการใช้แบบจำลอง 6 แบบ ได้แก่ Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LGBM) และ K-Nearest Neighbor (KNN) การทดลองทั้งหมดดำเนินการโดยใช้ไลบรารี scikit-learn ในภาษา Python ด้วยค่าพารามิเตอร์เริ่มต้น สำหรับการ

แบ่งข้อมูลตามลำดับ (Sequential Split) ผู้วิจัยได้ปรับแต่งโค้ด "continuous.py" จาก GitHub [<https://github.com/AndrzejTunkiel/USROP>] ให้เหมาะกับการวิจัย สำหรับ Continuous Learning ได้ปรับแต่งเพียงจำนวนการเพิ่ม Increments ขึ้นของข้อมูลเท่านั้น ในทางกลับกัน สำหรับ Sliding Window Split ได้ปรับแต่งการวนซ้ำ while เพื่อให้ชุดข้อมูลฝึกฝนในรอบถัดไป มาจากชุดข้อมูลทดสอบของรอบปัจจุบัน

3.5 วิธีการประเมินผลแบบจำลอง (Model Evaluation)

ในการวิจัยนี้เปรียบเทียบประสิทธิภาพแบบจำลองด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) ซึ่งต่างจากการวัดแบบอื่นๆ เช่น Mean Square Error (MSE) หรือ Root Mean Square Error (RMSE) ค่า MAE ให้การวัดความง่ายโดยตรงของค่าเฉลี่ยของข้อผิดพลาดในการทำนาย โดยไม่เน้นน้ำหนักไปที่ข้อผิดพลาดขนาดใหญ่ ในการวิจัยนี้กำหนดให้ค่าตัวแปรตามเป็นค่าทำนาย ROP หน่วยของอัตราการเจาะเป็น เมตรต่อชั่วโมง ดังนั้นค่า MAE จึงมีหน่วยเป็น m/hr และวัดค่าประสิทธิภาพแบบจำลองด้วยค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2) เพื่อใช้ในการวิเคราะห์ความแม่นยำของการทำนายผลของแบบจำลองแต่ละแบบ และสร้างกราฟเพื่อเปรียบเทียบค่า ROP จริงในอดีตกับค่าที่ทำนายได้จากแบบจำลอง

บทที่ 4

ผลการดำเนินงานวิจัย

ในการวิจัยนี้มีจุดประสงค์เพื่อศึกษาวิเคราะห์และทดลองเปรียบเทียบแบบจำลองเพื่อทำนายค่าอัตราการเจาะ (Rate of Penetration, ROP) โดยใช้ข้อมูลจากบริเวณ ไหล่ทวีปนอร์เวย์ (Norwegian Continental Shelf, NCS) อยู่ตอนกลางของทะเลเหนือ (Central part of North Sea) เป็นข้อมูลของบริษัทน้ำมัน Equinor ประเทศนอร์เวย์ ใช้การเรียนรู้ของเครื่อง (Machine Learning) โดยการแบ่งลำดับของข้อมูล (Sequential Data) เป็น 2 แบบ คือ การเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ผู้วิจัยได้ดำเนินการวิจัยโดยการศึกษาขั้นตอนต่างๆตลอดจนวัดประสิทธิภาพ เพื่อให้บรรลุจุดประสงค์ของการวิจัยที่ได้กำหนดไว้ ดังนี้

4.1 ผลลัพธ์การเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) ของการแบบจำลองต่างๆ

4.2 ผลลัพธ์การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ของการแบบจำลองต่างๆ

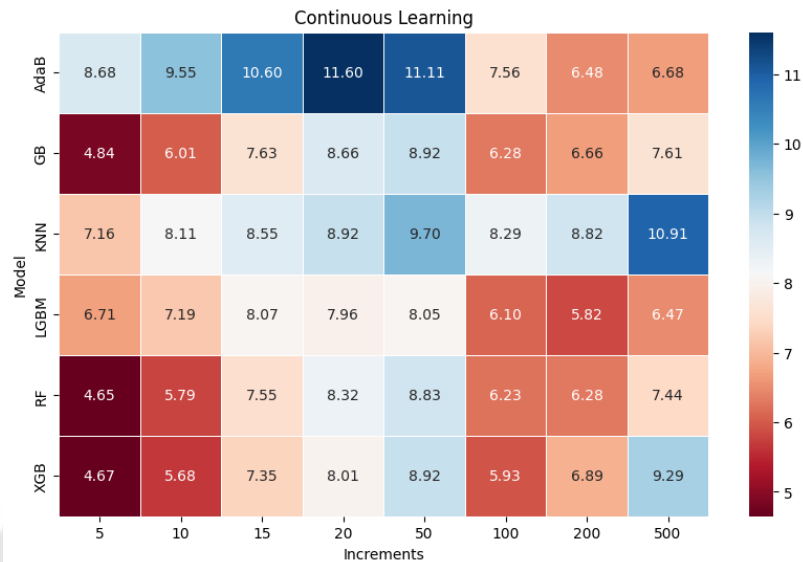
4.3 ผลลัพธ์เปรียบเทียบประสิทธิภาพการเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยค่าค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) ของแต่ละการเพิ่มขนาดข้อมูล (Increments)

4.4 ผลลัพธ์แบบจำลองในการทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะจริง

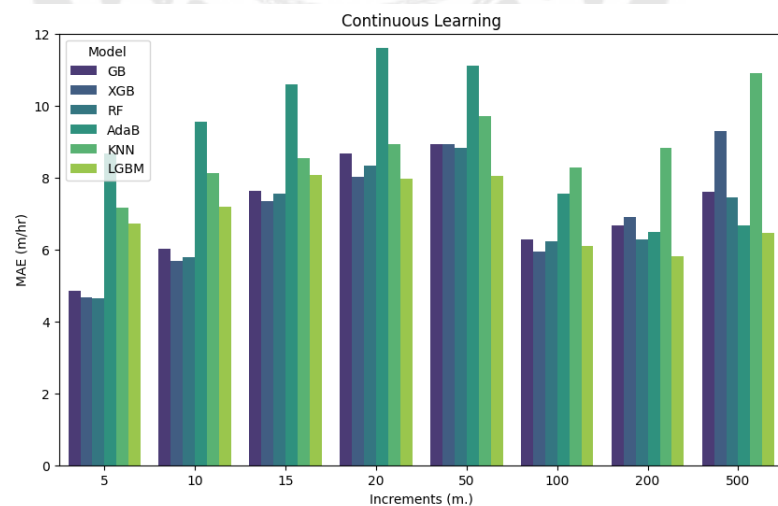
4.1 ผลลัพธ์การเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) ของการแบบจำลองต่างๆ

ในการแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง การเพิ่มขนาด increments เป็น 5, 10, 15, 20, 50, 100, 200 และ 500 เมตร ในแบบจำลองต่างๆ ได้แก่ Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), K-

Nearest Neighbor (KNN) และ Light Gradient Boosting Machine (LGBM) โดยการวัดประสิทธิภาพด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE)



ภาพประกอบ 67 แสดงค่า MAE ของการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) ใน แบบจำลองต่างๆ



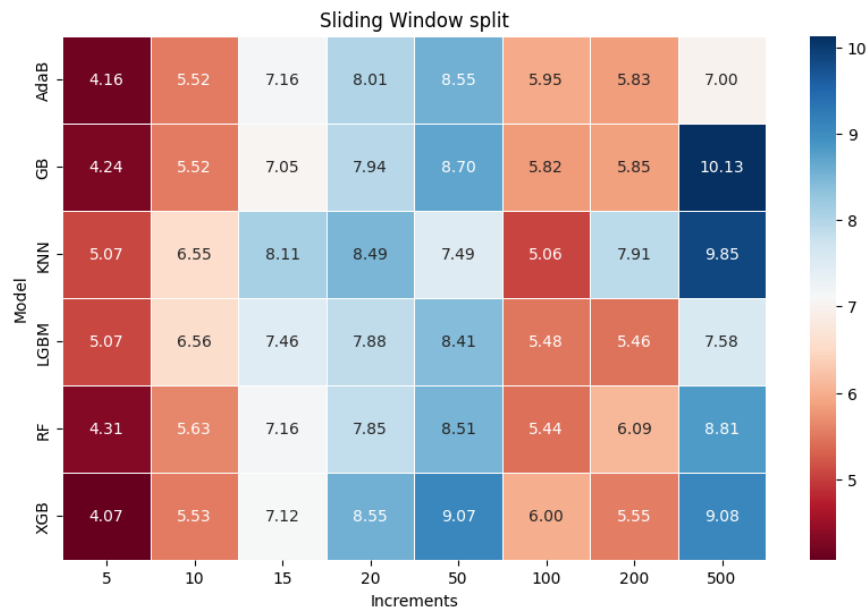
ภาพประกอบ 68 แสดงกราฟแท่งการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) แต่ละ increments ในการเปรียบเทียบแบบจำลองต่างๆกับ

ค่า MAE

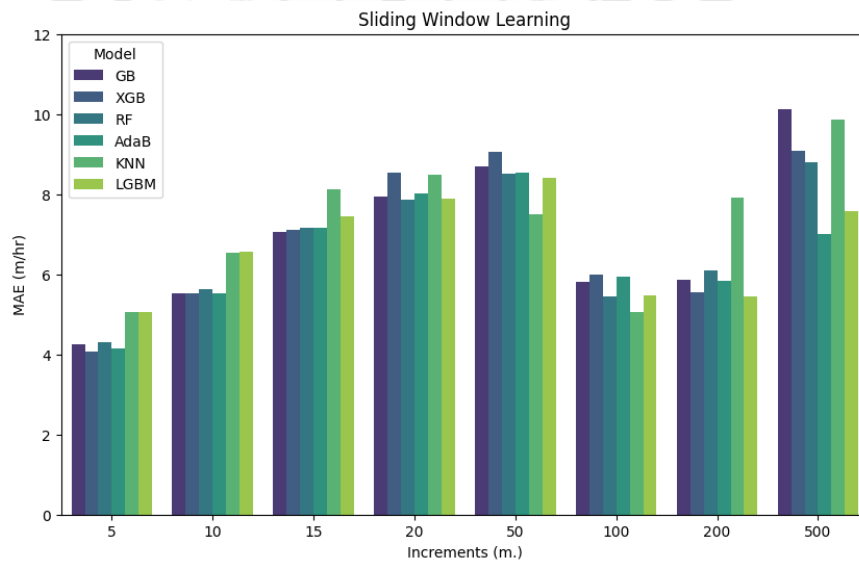
จากภาพประกอบที่ 67 ในตารางระหว่างการเพิ่ม increments ของข้อมูล กับแบบจำลองต่างๆ จะพบว่าค่าการเพิ่ม increments เท่ากับ 5 จะให้ค่าประสิทธิภาพด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์น้อยที่สุด เมื่อเปรียบเทียบกับค่าการเพิ่ม increments อื่นๆ โดยค่า MAE ที่น้อยที่สุดในแบบจำลอง Random Forest (RF) และ Extreme Gradient Boosting (XGB) ที่ใกล้เคียงกัน คือ 4.65 และ 4.67 m/hr ตามลำดับ โดยเมื่อทำการเปรียบเทียบด้วยกราฟแท่ง ดังภาพประกอบที่ 68 จะเห็นว่ากราฟแท่งที่ให้ MAE น้อยสุดจะอยู่ในการเพิ่ม Increments ข้อมูลที่ 5 เมตร เมื่อนำแบบจำลองมาเปรียบเทียบจะให้แบบจำลอง Random Forest (RF) ให้ประสิทธิภาพดีสุดในการเรียนรู้แบบต่อเนื่อง

4.2 ผลลัพธ์การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ของการแบบจำลองต่างๆ

ในการแบ่งข้อมูลการเรียนรู้แบบหน้าต่างบานเลื่อน การเพิ่มขนาดของข้อมูล (increments) เป็น 5, 10, 15, 20, 50, 100, 200 และ 500 เมตร ในแบบจำลองต่างๆ ได้แก่ Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), K-Nearest Neighbor (KNN) และ Light Gradient Boosting Machine (LGBM) โดยการวัดประสิทธิภาพด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE)



ภาพประกอบ 69 แสดงค่า MAE ของการเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ใน แบบจำลองต่างๆ

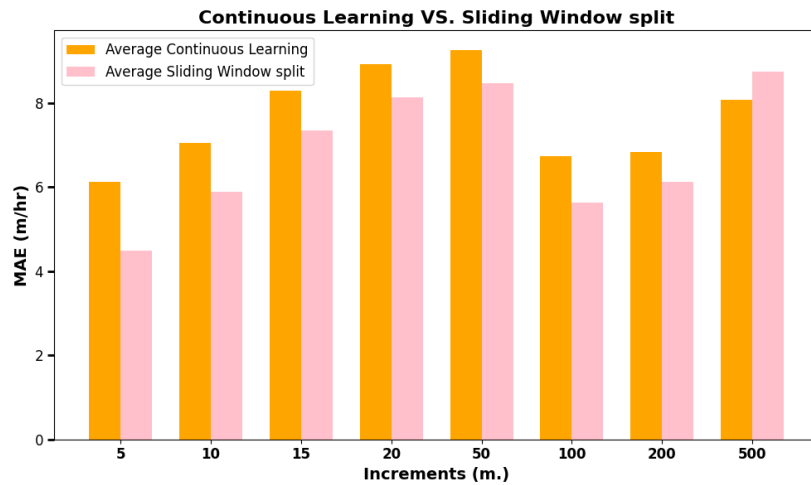


ภาพประกอบ 70 แสดงกราฟแท่งการเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) แต่ละ increments ในการเปรียบเทียบแบบจำลองต่างๆกับค่า MAE

จากภาพประกอบที่ 69 ในตารางระหว่างการเพิ่มขนาด (increments) ของข้อมูล กับแบบจำลองต่างๆ (Models) จะพบว่าค่าการเพิ่ม increments เท่ากับ 5 จะให้ค่าประสิทธิภาพด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) น้อยที่สุด เมื่อเปรียบเทียบกับค่าการเพิ่ม increments อื่นๆ โดยค่า MAE ที่น้อยที่สุดในแบบจำลอง Extreme Gradient Boosting (XGB) และ Adaptive Boosting (AdaBoost) คือ 4.07 และ 4.16 m/hr ตามลำดับ โดยเมื่อทำการเปรียบเทียบด้วยกราฟแท่ง ดังภาพประกอบที่ 70 จะเห็นว่ากราฟแท่งที่ให้ค่า MAE น้อยสุดจะอยู่ในการเพิ่ม increments ข้อมูลเท่ากับ 5 เมตร เมื่อนำแบบจำลองมาเปรียบเทียบจะให้แบบจำลอง XGB ให้ประสิทธิภาพดีสุดในการเรียนรู้แบบหน้าต่างบานเลื่อน

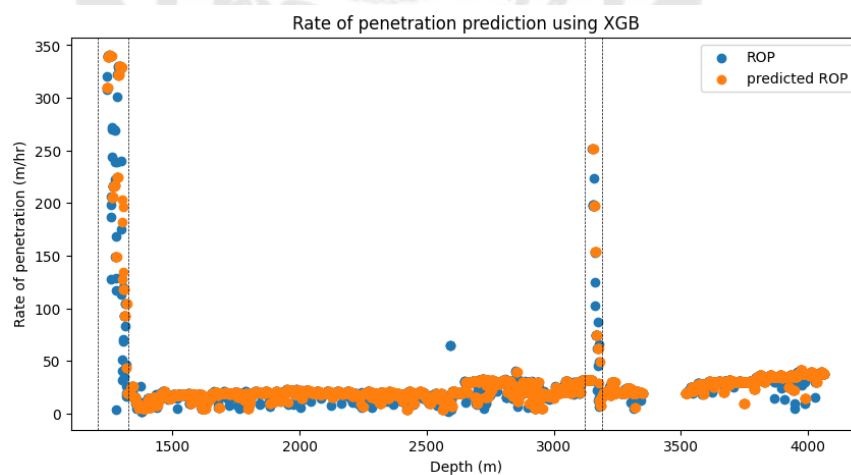
4.3 ผลลัพธ์เปรียบเทียบประสิทธิภาพการเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยค่า MAE ของแต่ละ Increments

จากภาพประกอบที่ 71 ข้อมูลทดสอบ (Testing) การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน มีประสิทธิภาพที่ดีกว่าในการวัดค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) เมื่อเทียบกับ การเรียนรู้แบบต่อเนื่อง ยกเว้นการเพิ่ม increments ของข้อมูลที่ 500 ที่สูงกว่าเล็กน้อย ผลลัพธ์แสดงให้เห็นว่า เมื่อขนาดการเพิ่มขึ้นของ increments แสดงค่า MAE ก็เพิ่มขึ้นเช่นกัน จนกระทั่งขนาดการเพิ่มขึ้นของ increments ถึงเท่ากับ 50 จากนั้นจะลดลงอีกครั้งที่ 100 ก่อนที่จะเพิ่มขึ้นอีกครั้ง ในการวัดประสิทธิภาพของแบบจำลองแบบผสมอย่าง Extreme Gradient Boosting (XGB), Random forest (RF) และ Gradient Boosting (GB) แสดงค่า MAE ที่น้อยเหมือนกันกับขนาดการเพิ่มขึ้นของ increments เช่น 5 และ 10 เช่นเดียวกับขนาดการเพิ่มขึ้นของ increments เช่น 100 และ 200 อย่างไรก็ตาม Light Gradient Boosting Machine (LGBM) และ Adaptive Boosting (AdaBoost) แสดงค่า MAE ที่น้อยกว่าในขนาดการเพิ่มขึ้นของ increments เช่น 100, 200 และ 500



ภาพประกอบ 71 แสดงการเปรียบเทียบระหว่างประสิทธิภาพการเรียนรู้อย่างต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยค่า MAE ของแต่ละ Increments

4.4 ผลลัพธ์แบบจำลองในการทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะจริง



ภาพประกอบ 72 แสดงการทำนายทำนายค่าอัตราการเจาะเปรียบเทียบกับค่าอัตราการเจาะจริงในแบบจำลอง XGB

จากภาพประกอบที่ 69 แสดงแบบจำลอง Extreme Gradient Boosting (XGB) ที่ใช้การแบ่งข้อมูลแบบหน้าต่างบานเลื่อนทีละ 5 หน่วย ได้ค่า MAE ที่น้อยที่สุดบนข้อมูลทดสอบ ประสิทธิภาพที่เหมาะสมของค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2) ของแบบจำลองนี้เท่ากับ 0.88 จากภาพประกอบที่ 72 เปรียบเทียบค่าอัตราการเจาะจริงกับค่าทำนายอัตราการเจาะโดยใช้กราฟแบบจุดกระจาย (Scatter plot) จุดสีน้ำเงินแทนค่า ROP จริง จากชุดข้อมูลทั้งหมด ในขณะที่จุดสีส้มแทนค่า ROP ที่ทำนาย เนื่องจากลักษณะของการแบ่งข้อมูลแบบหน้าต่างบานเลื่อนทีละ 5 หน่วย มีการฝึกฝน ทดสอบ และทำนายทั้งหมด 533 รอบ โดยเก็บเฉพาะค่าที่ทำนายเท่านั้น สังเกตว่าค่าที่ทำนายนั้นใกล้เคียงกับค่าจริง แต่มีความคลาดเคลื่อนบ้างช่วงที่เบี่ยงเบนไปจากการทำนาย ค่าความคลาดเคลื่อน (Residuals) ถูกคำนวณเป็นความแตกต่างระหว่างค่าจริงกับค่าที่ทำนาย เส้นประในภาพประกอบที่ 72 แสดงช่วงที่ค่าความคลาดเคลื่อนอยู่นอกเหนือช่วงที่ยอมรับได้ ซึ่งกำหนดไว้ที่ 3 เท่าของค่าเบี่ยงมาตรฐาน (Standard deviation) จากค่าความคลาดเคลื่อนเฉลี่ย (The average of the residuals)

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในการวิจัยเพื่อศึกษาวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตราการเจาะปิโตรเลียม ผู้วิจัยได้ประเมินประสิทธิภาพของแบบจำลองแต่ละอัลกอริทึม เพื่อนำมาเปรียบเทียบและสรุปผล จากผลที่ได้ดำเนินการวิจัยแล้ว สามารถแบ่งหัวข้อของการสรุปผลได้ดังนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในการวิจัยนี้ใช้ข้อมูลการเจาะแบบหลุมสำรวจ (Exploration Well) ชื่อชุดข้อมูล Volve oil field ชื่อหลุม 15/9-F15, StatoilHydro (Norway-StatoilHydro-15_47_9-F-15) มีข้อมูลทั้งหมด 2,701 แถว และ 10 คอลัมน์ ปัจจัยที่เกี่ยวข้อง ได้แก่

1. อัตราการเจาะ (Rate of Penetration, ROP)
2. ความลึกที่วัดได้ (Measure Depth, MD)
3. น้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB)
4. อัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM)
5. แรงบิดพื้นผิว (Surface Torque, T)
6. แรงดันในท่อ (Standpipe Pressure, SPP)
7. อัตราการไหลเข้าของน้ำโคลน (Mud Flow In, Q)
8. น้ำหนักของน้ำโคลนที่ออก (Mud Weight out, MW_{out})
9. เส้นผ่านศูนย์กลางของหัวเจาะ (Bit Diameter, D_{BIT})
10. พื้นที่หน้าตัดการไหลทั้งหมด (Total Flow Area, TFA)

เพื่อศึกษาวิเคราะห์เปรียบเทียบแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) โดยใช้แบบจำลอง Random Forest (RF), K Nearest Neighbor (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaB) และ Light Gradient Boosting Machine (LGBM) เพื่อทำนายอัตราการเจาะปิโตรเลียม (Rate of Penetration, ROP) โดยใช้เทคนิคการแบ่งข้อมูลแบบตามลำดับ (Sequential data) โดยอิงจากข้อมูลตามความลึก ประกอบด้วย การแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ Forward-Chaining Split) และการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) ด้วยการเพิ่มขนาด (Increments) ของข้อมูล 5, 10, 15, 20, 50, 100, 200, และ 500 เมตร เปรียบเทียบประสิทธิภาพแบบจำลองด้วยค่าผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Error, MAE) และวัดค่าประสิทธิภาพแบบจำลองด้วยค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2)

การวิจัยนี้พบว่าแบบจำลองที่ใช้เทคนิค Sliding Window มีประสิทธิภาพดีกว่า Continuous Learning ในข้อมูลชุดนี้ เนื่องจากการค่อยๆ เพิ่มค่า increment ในการฝึกฝนชุดข้อมูลของ Continuous Learning นั้น หมายถึง การเพิ่มค่าความลึกที่เพิ่มขึ้นเรื่อยๆ ยิ่งความลึกที่เพิ่มขึ้นจะทำให้ความเป็นเนื้อเดียวกัน (homogeneous) ของคุณลักษณะของชั้นหินลดลง ส่งผลให้พารามิเตอร์มีความแตกต่างกัน ในทางกลับกัน เทคนิค Sliding Window ข้อมูลชุดการฝึกฝน ขยับไปข้างหน้า ซึ่งยังคงความเป็นเนื้อเดียวกันของคุณลักษณะของชั้นหิน ทำให้ค่าพารามิเตอร์ ไม่ได้แตกต่างกันมากนัก จึงส่งผลให้ชุดข้อมูลของการทดสอบใน Sliding Window มีประสิทธิภาพที่ดีกว่า ดังภาพประกอบที่ 65 และ 66 ในการทดสอบเปรียบเทียบ แบบจำลอง XGB จึงให้ค่า MAE ต่ำที่สุด คือ 4.07 m/hr ด้วยเทคนิคการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) ที่ใช้การเพิ่ม increments ของข้อมูลที่ 5 เมตร และ วัดค่าประสิทธิภาพแบบจำลอง R^2 มีค่าเท่ากับ 0.88

5.2 อภิปรายผลการวิจัย

แม้ว่าแบบจำลองแบบเดิมจะมีประโยชน์ แต่ยังคงต้องการการประเมินค่าคงที่ต่างๆ ที่มาจากภาคสนามหรือวิธีการทดลอง ซึ่งทำให้ใช้เวลานานและการสรุปผลที่ยังจำกัดสำหรับทั้งหมด (Chandrasekaran & Kumar, 2020) การเรียนรู้ของเครื่องและโครงสร้างข้อมูลขนาดใหญ่ สามารถพัฒนาแบบจำลองแบบข้อมูลที่สามารทำนายค่าอัตราการเจาะที่ให้ความถูกต้องและมี

ประสิทธิภาพ ผลการวิจัยชี้ให้เห็นว่าสำหรับข้อมูลชุดนี้ เทคนิคการแบ่งหน้าต่างบานเลื่อน (Sliding Window split) ด้วยการเพิ่มขนาด (increments) ของข้อมูลเล็กๆ และแบบจำลองแบบรวมกลุ่ม (Ensemble models) ทำให้ประสิทธิภาพแบบจำลองที่ดีที่สุดในการทำนายค่าอัตราการเจาะ ในการเพิ่ม increments ของข้อมูลขนาดเล็กล้วนๆ ได้เปรียบเมื่อมีข้อมูลการฝึกฝนจำกัด แต่ต้องใช้การวนซ้ำ (iteration) หลายครั้งเพื่อฝึกฝนแบบจำลอง

เมื่อมีข้อมูลเพียงพอ การเพิ่มขึ้นของ increments สามารถลดจำนวนการวนซ้ำที่จำเป็นสำหรับชุดข้อมูลนี้ คุณสมบัติของชั้นหิน (Formation properties) (Ashrafi et al., 2019) (Gan et al., 2019) เป็นปัจจัยสำคัญในการบ่งบอกการเพิ่มขึ้นของ increments การเพิ่มขนาดของข้อมูลเล็กๆ หมายความว่า ข้อมูลยังคงอยู่ในชั้นหินเดียวกัน มีการศึกษาจำนวนมากที่แสดงให้เห็นว่าแบบจำลองแบบรวมกลุ่ม (Ensemble models) เช่น Extreme Gradient Boosting (XGB), Random Forest (RF), Gradient Boosting (GB) และ K-Nearest Neighbor (KNN) มักจะทำงานได้ดีในการทำนายค่าอัตราการเจาะ

การยอมรับได้ของค่าความคลาดเคลื่อน คำนวณจากความแตกต่างระหว่างค่าทำนายและค่าจริง ดังภาพประกอบที่ 73 กำหนดไว้ที่ 3 เท่าของค่าเบี่ยงเบนมาตรฐานจากค่าเฉลี่ยความคลาดเคลื่อน แสดงดังสมการที่ (11) และ (12)

กำหนดให้

$$\text{residuals} = \hat{y}_i - y_i \quad (11)$$

\bar{X} คือ ค่าเฉลี่ยของ residual

S_x คือ ส่วนเบี่ยงเบนมาตรฐานของ residual

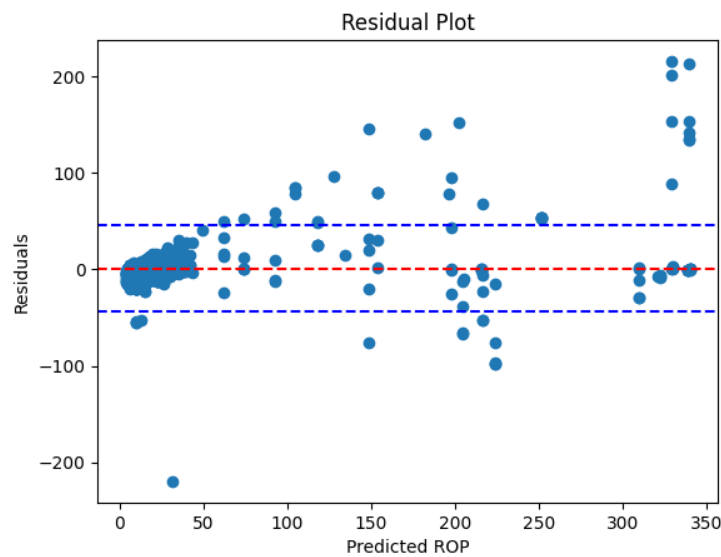
ค่าความคลาดเคลื่อนที่ยอมรับได้

$$\bar{X} - 3S_x \leq \text{residual} \leq \bar{X} + 3S_x \quad (12)$$

โดยที่

\hat{y} คือ ค่าการทำนาย

y_i คือ ค่าจริง



ภาพประกอบ 73 ค่าความคลาดเคลื่อนที่ยอมรับได้ไม่เกิน 3 เท่าของส่วนเบี่ยงเบน
มาตรฐาน

ในการวิจัยนี้ increment เท่ากับ 5 ได้ประสิทธิภาพดีที่สุดที่จุดนั้น โดยการทำให้ regression มักจะใช้ข้อมูลที่อยู่ในรูปแบบ (X,y) โดยที่ X คือค่า input variable และ y คือค่า output variable ในกำหนดให้ increments คือ n และ parameters คือ m ในบางกรณีการสร้าง regression model จะประเมินความเหมาะสมของข้อมูลที่มีอยู่ ซึ่งจำนวน n ไม่จำเป็นต้องเท่ากับจำนวน m ใน regression model สามารถใช้จำนวน n มากกว่าหรือน้อยกว่า m ก็ได้ ถ้าจำนวน n มากกว่า m อาจจะใช้โมเดลที่มีความซับซ้อนเพื่อการประมาณผลลัพธ์ที่แม่นยำมากขึ้นหรืออาจจะใช้ regularization techniques เพื่อควบคุมการ overfitting ของโมเดล ถ้ากำหนดให้ increment (n) เท่ากับ 5 และ parameters (m) เท่ากับ 10 สามารถใช้วิธีการ optimization algorithm เช่น gradient descent ดังนั้นการหา linear regression แสดงดังสมการที่ (13) เพื่อหาค่าของ parameters โดยใช้สูตร Normal Equation แสดงดังสมการที่ (14)

Linear regression model :

$$y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \dots + \theta_n X_n \quad (13)$$

โดยที่

X_1 ถึง X_n คือ ตัวแปรที่ส่งผลกระทบต่อความลึกในการขุดเจาะ

θ_0 ถึง θ_n คือ ค่าพารามิเตอร์ที่ต้องการหา

การหาค่าพารามิเตอร์ θ ที่เหมาะสมที่สุด :

$$\theta = (X^T X)^{-1} X^T y \quad (14)$$

โดยที่

θ = เวกเตอร์ของ parameters

X = เมทริกซ์ของ features ($n \times m$)

y = เวกเตอร์ของ target value ($n \times 1$)

การคำนวณนี้สามารถหาค่าของพารามิเตอร์ θ ที่ทำให้ regression model มีความเข้า กับข้อมูลที่ดีที่สุด จากนั้นนำค่า θ นี้มาใช้ในการทำนายค่า y สำหรับข้อมูลใหม่ที่ไม่เคยเห็นมา ก่อน

การศึกษานี้เน้นย้ำถึงความสำคัญของการพิจารณาข้อมูลตามความลึกและการนำไป รวมไว้ในแบบจำลองของการทำนายการขุดเจาะหลุมสำรวจแบบเวลาจริง (Real Time)

5.3 ข้อเสนอแนะ

จากปัญหาที่พบในการดำเนินงานวิจัยนี้ ผู้วิจัยได้สังเกตเห็นว่าขั้นตอนการแบ่งข้อมูลสำหรับฝึกฝน (Training) และทดสอบ (Testing) มีความสำคัญมาก เมื่อมีข้อมูลการฝึกฝนที่จำกัด และปัจจัยที่ส่งผลต่อการเพิ่มขึ้นของขนาดข้อมูล คือ คุณสมบัติของชั้นหิน (Formation properties) ที่จะนำมาพิจารณาเกี่ยวข้องกับการทำนายค่าอัตราการเจาะในอนาคต รวมถึง วิธีการในการทำ cross validation และการปรับแต่งพารามิเตอร์ (Hyperparameter Tuning) ในแต่ละรอบการฝึกฝน



บรรณานุกรม

- Ashrafi, S. B., Anemangely, M., Sabah, M., & Ameri, M. J. (2019). Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. *Journal of Petroleum Science and Engineering*, 175, 604–623.
<https://doi.org/10.1016/j.petrol.2018.12.013>
- Bingham, M. G. (1964). A New Approach to Interpreting Rock Drillability. *Oil Gas Journal*, 62.
- Bourgoyne, A. T., Millheim, K. K., Chenevert, M. E., & Young, F. S. (1986). *Applied drilling engineering* (Vol. 2). Society of Petroleum Engineers Richardson.
- Bourgoyne, A. T., & Young, F. S. (1974). A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection. *Society of Petroleum Engineers Journal*, 14(04), 371–384. <https://doi.org/10.2118/4238-PA>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Chandrasekaran, S., & Kumar, G. S. (2020). Drilling efficiency improvement and rate of penetration optimization by machine learning and data analytics. *International Journal of Mathematical, Engineering and Management Sciences*, 5(3), 381–394.
<https://doi.org/10.33889/IJMEMS.2020.5.3.032>
- Dhuha T., A.-S., Allawi, R. H., Al-Mudhafar, W. J., & Yao, C. (2023). *Ensemble Machine Learning for Data-Driven Predictive Analytics of Drilling Rate of Penetration (ROP) Modeling: A Case Study in a Southern Iraqi Oil Field*. SPE-213043-MS.
- El-Sayed, Y., Salem, A. M., & El-Rammah, S. (2023). Journal of Mining and Environment (JME) Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms. *Journal of Mining and Environment*

- (*JME*), 14(3), 755–770. <https://doi.org/10.22044/jme.2023.12694.2306>
- Fix, E., & Hodges, J. L. (1951). *Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine.
<https://books.google.co.th/books?id=4XwytAEACAAJ>
- Gan, C., Cao, W. H., Wu, M., Chen, X., Hu, Y. L., Liu, K. Z., Wang, F. W., & Zhang, S. B. (2019). Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China. *Journal of Petroleum Science and Engineering*, 181. <https://doi.org/10.1016/j.petrol.2019.106200>
- Hareland, G., & Rampersad, P. R. (1994). Drag-bit model including wear. *SPE Latin America and Caribbean Petroleum Engineering Conference*, SPE-26957.
- Hegde, C., Daigle, H., Millwater, H., & Gray, K. (2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *Journal of Petroleum Science and Engineering*, 159, 295–306. <https://doi.org/10.1016/j.petrol.2017.09.020>
- Maurer, W. C. (1962). The “Perfect—Cleaning” Theory of Rotary Drilling. *Journal of Petroleum Technology*, 14(11), 1270–1274. <https://doi.org/10.2118/408-PA>
- Motahhari, H. R., Hareland, G., & James, J. A. (2010). Improved Drilling Efficiency Technique Using Integrated PDM and PDC Bit Parameters. *Journal of Canadian Petroleum Technology*, 49(10), 45–52. <https://doi.org/10.2118/141651-PA>
- Murray, A. S., & Cunningham, R. A. (1955). Effect of Mud Column Pressure on Drilling Rates. *Transactions of the AIME*, 204(01), 196–204. <https://doi.org/10.2118/505-G>
- Olafadehan, O. A., & Ahaotu, I. D. (2023). Comparative Analysis of Machine Learning Algorithms in Predicting Rate of Penetration during Drilling. In *J Petro Chem Eng* (Vol. 1, Issue 1, pp. 32–47). <https://urfpublishers.com/journal/petrochemical-engineering>
- Ren, Y., Lu, B., Zheng, S., Bai, K., Cheng, L., Yan, H., & Wang, G. (2023). Research on the Rate of

- Penetration Prediction Method Based on Stacking Ensemble Learning. *Geofluids*, 2023.
<https://doi.org/10.1155/2023/6645604>
- Soares, C., Daigle, H., & Gray, K. (2016). Evaluation of PDC bit ROP models and the effect of rock strength on model coefficients. *Journal of Natural Gas Science and Engineering*, 34, 1225–1236. <https://doi.org/10.1016/j.jngse.2016.08.012>
- Soares, C., & Gray, K. (2019). Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *Journal of Petroleum Science and Engineering*, 172, 934–959. <https://doi.org/10.1016/j.petrol.2018.08.083>
- Speer, J. W. (1959). A Method for Determining Optimum Drilling Techniques. *SPE Upper Gulf Coast Drilling and Production Conference*, SPE-1242-G. <https://doi.org/10.2118/1242-G>
- T. Cover & P. Hart. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
- Tin Kam Ho. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1, 278–282 vol.1.
<https://doi.org/10.1109/ICDAR.1995.598994>
- Tunkiel, A. T., Sui, D., & Wiktorski, T. (2021). Reference dataset for rate of penetration benchmarking. *Journal of Petroleum Science and Engineering*, 196.
<https://doi.org/10.1016/j.petrol.2020.108069>
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>



ประวัติผู้เขียน

