

การวิเคราะห์เปรียบเทียบแบบจำลองโดยการเรียนรู้ของเครื่องเพื่อทำนายอัตราการขาดเจาะปิโตรเลียม

ญาณเดช ภาควัฒนา¹, ศุภร คนธภักดี²

บทคัดย่อ

การดำเนินงานการขาดเจาะปิโตรเลียม ปัจจัยหนึ่งที่มีผลโดยตรงต่อการขาดเจาะ คือ ค่าอัตราการเจาะ (Rate of Penetration, ROP) การทำนายค่าอัตราการเจาะที่ถูกต้องเป็นปัจจัยสำคัญที่จะส่งผลต่อการควบคุมค่าใช้จ่ายในการดำเนินงานการขาดเจาะ งานวิจัยนี้เป็นการศึกษาการเรียนรู้ของเครื่องในรูปแบบจำลองการเรียนรู้แบบรวมกลุ่ม เช่น Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LGBM) และ K-Nearest Neighbor (KNN) ข้อมูลที่ใช้ในการทำนายค่าอัตราการเจาะนำมาจากบริษัท Equinor บริเวณตอนกลางทะเลเหนือ ประเทศนอร์เวย์ โดยชุดข้อมูลชื่อว่า Volve การวิจัยนี้เป็นกาจำลองเหมือนสถานการณ์ของความเป็นจริง โดยเฉพาะในกรณีที่มีข้อจำกัดทางด้านข้อมูล ผู้วิจัยใช้การลำดับของข้อมูล (Sequential data) หรือข้อมูลความลึกที่ต่อเนื่องกัน (Depth-dependent data) ทำการฝึกฝนและทดสอบรูปแบบจำลองการเรียนรู้ของเครื่องแบบต่างๆ โดยการศึกษาด้วยเทคนิคการเรียนรู้อย่างต่อเนื่อง (Continuous Learning) และ การเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Learning) และวัดประสิทธิภาพของรูปแบบจำลองด้วยค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ผลของการวิจัยนี้ให้แบบจำลองการเรียนรู้แบบรวมกลุ่มที่เป็น Extreme Gradient Boosting (XGB) ด้วยเทคนิคการเรียนรู้แบบหน้าต่างบานเลื่อน (Sliding Window Learning) ในการทำนายค่าอัตราการเจาะที่มีประสิทธิภาพดีที่สุด อย่างไรก็ตาม ความแตกต่างของขนาดข้อมูลของการเรียนรู้แบบหน้าต่างบานเลื่อนยังให้ผลการศึกษานาของข้อมูลที่เหมาะสมที่สุด

คำสำคัญ : อัตราการเจาะ, การดำเนินงานการเจาะ, รูปแบบการเรียนรู้ของเครื่อง, แบบจำลองการเรียนรู้แบบรวมกลุ่ม, การเรียนรู้อย่างต่อเนื่อง, การเรียนรู้แบบหน้าต่างบานเลื่อน

¹ หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

² คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author: Tel.: 094-9491954 E-mail address: yanadade.pak@swu.ac.th

Comparative Analysis of Machine Learning Algorithms for Drilling Rate Prediction

Yanadade Pakawatthapana^{1*}, Subhorn Khonthapagdee²

Abstract

In drilling operations, the Rate of Penetration directly impacts drilling efficiency. Accurately predicting 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, Light Gradient Boosting Machine and K-Nearest Neighbor were used to predict ROP using a Volve dataset from the Norwegian North Sea provided by Equinor Company. 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 various machine learning models using sequential split such as continuous learning and sliding window split techniques. The model's performances 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.

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

¹ Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

² Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author: Tel.: 094-9491954 E-mail address: yanadade.pak@g.swu.ac.th

1. บทนำ

อุตสาหกรรมปิโตรเลียมที่เกี่ยวข้องกับการดำเนินงานการขุดเจาะ สิ่งที่ต้องคำนึงถึงเป็นการเพิ่มประสิทธิภาพและควบคุมค่าใช้จ่ายในการขุดเจาะ ปัจจัยสำคัญที่ส่งผลโดยตรงกับการขุดเจาะ คือ ค่าอัตราการเจาะ (Rate of Penetration, ROP) เป็นการวัดความเร็วของหัวเจาะขณะเจาะเข้าไปในชั้นหิน การทำนายค่าอัตราการเจาะซึ่งเป็นปัจจัยที่สำคัญในการดำเนินงานการขุดเจาะ แต่ยังมีปัจจัยต่างๆที่เกี่ยวข้อง เช่น คุณสมบัติของของไหลในการเจาะ (Drilling fluid properties) เช่น ค่าความหนืดของของไหล (Viscosity) และ ค่าความหนาแน่น (Density) และลักษณะเฉพาะของชั้นหิน (Formation characteristics) รวมถึง น้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB), อัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM), ชนิดและรูปแบบของหัวเจาะ (Bit type and Design) และ คุณสมบัติของชั้นหิน (Formation properties) วิธีการทำนายค่าอัตราการเจาะ (Rate of Penetration, ROP) แบบเดิมจะใช้แบบจำลองเชิงประจักษ์ (Empirical models) ซึ่งเป็นทฤษฎีแบบจำลองทางคณิตศาสตร์ที่มาจาก การคำนวณทางคณิตศาสตร์มีการเก็บสถิติไว้เป็นข้อมูลและความรู้เฉพาะทาง ปัจจุบันเทคนิคทางข้อมูล (Data-Driven) การเรียนรู้ของเครื่อง (Machine Learning) แบบจำลองการถดถอย (Regression models) ที่ให้ผลการทำนายค่าอัตราการเจาะที่สัมพันธ์กับปัจจัยที่เกี่ยวข้อง โดยการเรียนรู้ของเครื่อง (Machine Learning) เป็นที่ได้รับความสนใจในอุตสาหกรรมปิโตรเลียม ทางด้านการขุดเจาะเพื่อเพิ่มประสิทธิภาพที่ถูกต้องในการดำเนินงานการขุดเจาะ เป็นการเข้าถึงประโยชน์จากชุดข้อมูลขนาดใหญ่ที่มีความหลากหลายต่อปัจจัยที่มีผลต่อการเจาะ เช่น น้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB), อัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM), คุณสมบัติของของไหลในการเจาะ (Drilling Fluid properties), ชนิดของหัวเจาะ (Bit type) และลักษณะเฉพาะของชั้นหิน (Formation characteristics) ที่นำมาทำนายค่าอัตราการเจาะ โดยปกติแล้ว รูปแบบการเรียนรู้ของเครื่อง (Machine Learning) จะต้องอาศัยการฝึกฝน (Training) และการทดสอบ (Testing) กับข้อมูลการเจาะซึ่งนำมาถูกพัฒนาให้มีความถูกต้องกับแบบจำลองของการทำนายค่าอัตราการเจาะ อย่างไรก็ตาม การขุดเจาะหลุมสำรวจ (Exploration Well) ข้อมูลที่ได้จากชั้นหินใต้ดิน (Subsurface Formation) ในส่วนนี้ข้อมูลที่มีอยู่ยังมีข้อจำกัดเมื่อเริ่มทำการขุดเจาะ

การศึกษานี้มีวัตถุประสงค์ ในการเรียนแบบการทำนายค่าอัตราการเจาะแบบเวลาจริง เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost), Light Gradient Boosting (LGBM) และ K-Nearest Neighbor (KNN) ใช้ 2 เทคนิคของความแตกต่างในการฝึกฝน (Training) และการทดสอบ (Testing) ที่ขึ้นกับชุดข้อมูลความลึกที่ต่อเนื่องกันในการขุดเจาะ (Depth-dependent drilling dataset)

2. ความเป็นมาและงานวิจัยที่เกี่ยวข้อง

วิธีการทำนายค่าอัตราการเจาะแบบเดิมจะอาศัยแบบจำลองทางฟิสิกส์โดยใช้สูตรและฟังก์ชันทางคณิตศาสตร์ แบบจำลอง Speer [1] อยู่บนพื้นฐานทางคณิตศาสตร์ที่มีปัจจัยสำคัญในการขุดเจาะ คือ น้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB), อัตรารอบการหมุนของก้านเจาะ (Rotary Speed, RPM) และการขับเคลื่อนด้วยไฮดรอลิก (Hydraulic power) คล้ายคลึงกับแบบจำลอง Bingham [2] ที่พิจารณาชนิดของชั้นหิน (Formation type) รวมไปถึงแบบจำลอง Hareland & Rampersad [3] และ Mohatari และคณะ [4] ที่ยังพิจารณาถึงความแข็งของหิน (Rock strength) เข้าไปในแบบจำลองด้วย แม้ว่าแบบจำลองแบบเดิมจะมีประโยชน์ แต่ยังคงต้องการการประเมินค่าที่ต่างๆ ที่มาจากภาคสนามหรือวิธีการทดลอง ซึ่งทำให้ใช้เวลานานและการสรุปผลที่ยังจำกัดสำหรับทั้งหลุม [5] ในการพัฒนาล่าสุดในการเรียนรู้ของเครื่องและโครงสร้างข้อมูลขนาดใหญ่สามารถพัฒนาแบบจำลองข้อมูลที่สามารถทำนายค่าอัตราการเจาะที่ให้ความถูกต้องและมีประสิทธิภาพ ในการศึกษาของ Hegde และคณะ [6] แสดงแบบจำลอง Random Forest (RF) ที่ให้ค่าความถูกต้องและค่าการตัดสินใจ (R^2) ที่สูง เมื่อเปรียบเทียบกับแบบจำลองแบบเดิม

ทางฟิลิกส์ Soares & Gray [7] ได้ทำการศึกษาและสรุปว่า การเรียนรู้ของเครื่อง (Machine Learning) รวมถึง Random Forest (RF), Support Vector Machine (SVM) และ Neural Network มีประสิทธิภาพเหนือกว่าแบบจำลองแบบเดิม ดังนั้น รูปแบบการเรียนรู้ของเครื่องจึงเป็นแนวทางที่มีแนวโน้มที่ใช้ในการทำนายอัตราการเจาะในขณะขุดเจาะจากการศึกษาของ Ren และคณะ [8] การเรียนรู้ของเครื่องแบบรวมกลุ่ม (Ensemble machine learning) จะรวมแบบจำลองหลายแบบจำลองเพื่อสร้างแบบจำลองที่ใช้ในการทำนายที่มีความถูกต้องมากขึ้น การใช้จุดแข็งของแต่ละแบบจำลองเพื่อปรับปรุงประสิทธิภาพโดยรวม แบบจำลองแบบรวมกลุ่ม เช่น Random Forest (RF), Gradient Boosting (GB), Adaptive Boosting (AdaBoost) จะแสดงค่าความแม่นยำและค่าความถูกต้องที่สูงในการทำนายอัตราการเจาะจากการศึกษาของ Al-Sahlanee และคณะ [9], Hedge และคณะ [6] และ Soares & Gray [7] K-Nearest Neighbor (KNN) เป็น Non-parametric ทั้ง Classification และ Regression ต่างจากแบบจำลองแบบเดิมที่เป็นแบบจำลอง Parametric การใช้ K-Nearest Neighbor (KNN) เพื่อทำนายอัตราการเจาะได้การสนับสนุนจากการวิจัยหลายวิจัย รวมถึงการวิจัยของ El-Sayed และคณะ [10] และ Olafadehan & Ahaotu [11]

การพัฒนาแบบจำลองเพื่อทำนายความแม่นยำในการเรียนรู้ของเครื่องโดยขั้นตอนที่สำคัญจะเป็นการฝึกฝนและการทดสอบ ซึ่งจะเกี่ยวข้องกับการแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดการฝึกฝน (Training set) เพื่อใช้ในการฝึกฝนแบบจำลอง และ ชุดการทดสอบ (Testing set) เพื่อใช้ประเมินประสิทธิภาพของแบบจำลอง แนวทางการแบ่งข้อมูลโดยทั่วไปจะเป็นการแบ่งข้อมูลแบบสุ่ม (Random) โดย Hedge และคณะ [6] และ Tunkiel และคณะ [12] แนะนำให้แบ่งข้อมูลเป็นลำดับของข้อมูล (Sequential data) ที่ใช้ในการจำลองกระบวนการขุดเจาะ การแบ่งข้อมูลตามลำดับจะยังคงความสัมพันธ์ตามความลึกและรูปแบบกระบวนการเจาะเอาไว้ Tunkiel และคณะ [12] ได้แนะนำ การเรียนรู้แบบต่อเนื่อง (Continuous Learning) ที่ใช้ในการฝึกฝนและการทดสอบแบบจำลองทุกๆ 30 เมตร ข้อมูลที่แบ่งจะคล้ายกับการวิเคราะห์อนุกรมเวลา (Time Series Analysis) ที่แบ่งแบบ Forward-chaining หรือ Rolling-origin โดยชุดการฝึกฝน (Training set) และชุดการทดสอบ (Testing set) ที่ถูกแบ่งเป็นส่วนๆตามลำดับในแบบจำลองจะถูกฝึกฝนกับข้อมูลในอดีต (Training on past data) และ ทดสอบกับข้อมูลในอนาคต (Test on future data) โดยการศึกษาของ Tashman [13]

งานวิจัยนี้มีส่วนช่วยในการต่อยอดงานวิจัยของ Tunkiel และคณะ [12] ซึ่งมุ่งเน้นไปที่สถานการณ์การเรียนรู้แบบต่อเนื่อง (Continuous Learning) โดยผู้วิจัยได้ทำการวิเคราะห์และเปรียบเทียบประสิทธิภาพอัลกอริทึมการเรียนรู้ของเครื่องที่แตกต่างกัน โดยใช้เทคนิค 2 แบบในการแบ่งข้อมูลแบบลำดับ ได้แก่ การแบ่งข้อมูลการเรียนรู้แบบต่อเนื่อง (Continuous Learning หรือ forward-chaining split) และ การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split) นอกจากนี้ ผู้วิจัยยังได้ศึกษาขนาดของหน้าต่าง (Window size) ที่แตกต่างกันในการทำหน้าต่างบานเลื่อนที่จะส่งผลต่อความแม่นยำของแบบจำลองการทำนายอัตราการเจาะที่ต้องการ

3. วิธีดำเนินการ

3.1 การเตรียมข้อมูล (Data Preparation)

ขั้นตอนที่ 1 : แหล่งของข้อมูล (Data Source)

การศึกษานี้ใช้ข้อมูลจริงจากบริษัท Equinor ประเทศนอร์เวย์ ที่ได้จากการขุดเจาะกลางทะเลโดยดำเนินการอยู่ช่วงระหว่างปี 2008 ถึง 2016 [12] และเปิดเป็น open source ในปี 2018 ชุดข้อมูลที่ให้มีชื่อว่า Volve oil field ซึ่งเป็นข้อมูลบริเวณ

ไหล่ทวีปนอร์เวย์ (Norwegian Continental Shelf, NCS) อยู่ตอนกลางทะเลเหนือ (Central part of North Sea) ห่างจาก Stavanger ทางตะวันตกประมาณ 200 กิโลเมตร และอยู่ทางด้านเหนือของ Sleipner ตะวันออกประมาณ 5 กิโลเมตร

ขั้นตอนที่ 2 : การได้มาของข้อมูล (Data Acquisition)

ชุดข้อมูล Volve Field สามารถเข้าถึงผ่านทาง Equinor Open Data License ซึ่งสามารถดูรายละเอียดเพิ่มเติมได้ที่เว็บไซต์ของ Equinor [14] ผู้วิจัยศึกษามุ่งเน้นไปที่หลุมเจาะแบบสำรวจมีชื่อหลุม 15/9-F15, StatoilHydro (Norway-StatoilHydro-15_9-F15) ข้อมูลในการชุดเจาะแบ่งออกเป็น 2 ส่วน ได้แก่ 17.5” section และ 8.5” section ข้อมูลถูกเก็บใน Wellsite information transfer standard markup language (WITSML) นำเข้าและวิเคราะห์โดย Python ขั้นตอนการเข้าถึงข้อมูลและการนำเข้าข้อมูลด้วย Python สามารถศึกษาเพิ่มเติมได้ที่เว็บไซต์ DiscoverVolve [15] และ GitHub [16] มีข้อมูลทั้งหมด 112,714 แถว และ 189 คอลัมน์ และความลึกเริ่มจาก 1178.509 เมตร ถึง 4065.346 เมตร

ขั้นตอนที่ 3 : การทำความสะอาดและการเตรียมข้อมูล (Data Cleaning and Preprocessing)

ในหัวข้อนี้ผู้วิจัยจะให้รายละเอียดในการเตรียมและทำความสะอาดข้อมูล โดยผู้วิจัยได้ทำการลบค่าความลึกตั้งแต่ต้นจนถึงค่าความลึก 1200 เมตร เนื่องจากข้อมูลในช่วงดังกล่าวไม่ได้ถูกบันทึกไว้ ในการดำเนินการชุดเจาะ 17.5” section ทำการชุดเจาะถึงค่าความลึก 2591.141 เมตรและก่อนที่จะเจาะ 8.5” section ค่าความลึกจะเริ่มต้นที่ 2464.156 เมตร ซึ่งจะพบข้อมูลที่ซ้ำซ้อนกัน จึงได้ทำการลบข้อมูลดังกล่าวออก คอลัมน์ที่บ่งบอกขนาดของหัวเจาะเปลี่ยนชื่อคอลัมน์เป็น Bit Diameter ค่าที่ขาดหายไปจะใช้วิธีการเติมข้อมูลด้วย Forward Fill และ Backward Fill ค่าความลึกที่วัดได้จะทำการปัดเศษเป็นเลขศูนย์และทำการลบข้อมูลที่ซ้ำซ้อนออก โดยเก็บไว้เพียงค่าเริ่มต้นเพื่อแทนค่าความลึกนั้นๆ และทำการเพิ่มคอลัมน์ใหม่ชื่อ Total Flow Area โดยชุดข้อมูลทั้งหมดจะเหลือ 2,701 แถว และ 10 คอลัมน์

ขั้นตอนที่ 4 : การคัดเลือกคุณลักษณะ (Feature Selection)

ในการทำนายค่าอัตราการเจาะตัวแปรที่มีผลต่ออัตราการเจาะ ตัวอย่างเช่น น้ำหนักที่ตกลงบนหัวเจาะ (Weight on Bit, WOB), อัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM), คุณสมบัติของของไหลในการเจาะ (Drilling fluid properties), ชนิดของหัวเจาะ (Bit type) และ ลักษณะเฉพาะของชั้นหิน (Formation characteristics) โดยการศึกษาของ Ashrafi และคณะ [17], Gan และคณะ [18] และ Soares & Gray [7] คุณลักษณะเป็นสิ่งสำคัญที่มีผลโดยตรงต่อการดำเนินการชุดเจาะและส่งผลกระทบต่ออัตราการเจาะ การวิจัยนี้ คัดเลือกคุณลักษณะสำหรับแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ อัตราการเจาะ (Rate of Penetration, ROP), ความลึกที่วัดได้ (Measure Depth, MD), น้ำหนักที่ตกลงบนหัวเจาะ (Weight on Bit, WOB), อัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM), แรงบิดพื้นผิว (Surface Torque, T), แรงดันในท่อ (Standpipe Pressure, SPP), อัตราการไหลเข้าของน้ำโคลน (Mud Flow In, Q), น้ำหนักของน้ำโคลนที่ออก (Mud Weight out, MW_{out}), เส้นผ่านศูนย์กลางของหัวเจาะ (Bit Diameter, D_{BIT}) และ พื้นที่หน้าตัดการไหลทั้งหมด (Total Flow Area, TFA) คำศัพท์คุณลักษณะทางเทคนิคในการชุดเจาะปิโตรเลียมและความหมาย แสดงในตารางที่ 1

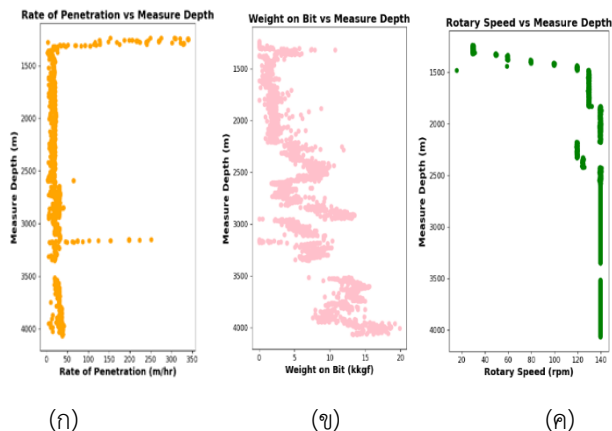
ตารางที่ 1 คำศัพท์คุณลักษณะทางเทคนิคในการขุดเจาะปิโตรเลียมและความหมาย

Feature	Definition	Unit
Rate of Penetration (ROP)	A measure of the speed at which the bit drill into formation	Meter / hour (m/hr)
Measure Depth (MD)	The length of the drilled borehole	Meter (m)
Weight on Bit (WOB)	The amount of downward force placed on the Bit	Kilo-kilogram force (kkgf)
Rotary Speed (RPM)	The speed at which the drill string and bit rotate during the drilling process	Revolution per minute (rpm)
Surface Torque (T)	The amount of rotational force or torque applied at the surface of the wellbore	Kilonewtons. Meter (kN.m)
Standpipe Pressure (SPP)	The pressure of the drilling fluid within the standpipe	Kilopascal (kPa)
Mud Flow In (Q)	Circulation of drilling mud into the wellbore during the drilling process	Liter / minute (L/min)
Mud Weight Out (MW _{out})	The process of measuring and adjusting the density or weight of the drilling fluid as it returns through the flowline	Grams / cubic centimeter (g/cm ³)
Bit Diameter (D _{BIT})	The diameter of the drill bit	Inch
Total Flow Area (TFA)	The set of nozzle areas which fluid flow through a bit	Inch ²

3.2 การวิเคราะห์ข้อมูลเชิงการตรวจสอบเบื้องต้น (Exploratory Data Analysis, EDA)

การหาข้อมูลเชิงลึกและทำความเข้าใจในความสัมพันธ์ระหว่างตัวแปรเป็นการวิเคราะห์ข้อมูลเชิงการตรวจสอบเบื้องต้น ขุดข้อมูลประกอบด้วย 2,701 แถว และ 10 คอลัมน์ ทำการลบข้อมูลที่มีค่าความผิดปกติของอัตราการเจาะมากกว่า 400 ออก ข้อมูลที่ได้ทั้งหมดมี 2,664 แถว และ 10 คอลัมน์ ที่ค่าความลึกระหว่าง 2,592 ถึง 2,604 เมตร พบว่าค่าน้ำหนักที่กดลงบนหัวเจาะ (Weight on Bit, WOB) มีค่าสูงผิดปกติเนื่องจากการเปลี่ยนแปลงของชั้นหิน โดยทำการลบและแก้ไขด้วยวิธีเชิงเส้น (Linear Interpolation) จากภาพที่ 1 (ก) เป็นกราฟที่พล็อตระหว่าง ค่าความลึกที่วัดได้ (Measure Depth, MD) กับ ค่าอัตราการเจาะ (Rate of Penetration, ROP) ในช่วงเริ่มต้นค่าอัตราการเจาะจะมีค่าสูงที่ค่าความลึกระหว่าง 1,200 ถึง 1,306 เมตร หลังจากนั้นค่าอัตราการเจาะจะค่อยๆลดลงและเริ่มคงที่ที่ค่าความลึก 1,320 เมตร ในส่วน 17.5”section ค่าอัตราการเจาะอยู่ระหว่าง 5 ถึง 25 ขณะที่ 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 ในภาพที่ 1 (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) ช่วง

นี้จะอยู่ที่ 6 หลังจากนั้นจะลดค่าน้ำหนักที่ตกลงบนหัวเจาะ (Weight on Bit, WOB) และจะเพิ่มค่าน้ำหนักที่ตกลงบนหัวเจาะให้อยู่ระหว่าง 7 ถึง 19 ค่าอัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM) จากภาพที่ 1 (ค) ใน 17.5” section ค่าอัตรารอบของการหมุนของก้านเจาะค่อยๆเพิ่มขึ้นจนถึงที่ค่าความลึก 1,480 เมตร หลังจากนั้นค่าอัตรารอบของการหมุนของก้านเจาะจะคงที่ที่ 130 และ 8.5” section ค่าอัตรารอบของการหมุนของก้านเจาะประมาณ 140



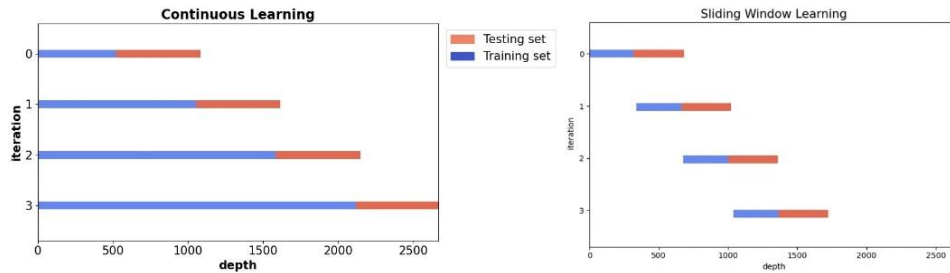
ภาพที่ 1 การอธิบายเปรียบเทียบปัจจัยกับช่วงค่าความลึกต่างๆที่วัดได้ (Measure Depth, MD) (ก) ค่าอัตราการเจาะ (Rate of Penetration, ROP) (ข) ค่าน้ำหนักที่ตกลงบนหัวเจาะ (Weight on Bit, WOB) (ค) ค่าอัตรารอบของการหมุนของก้านเจาะ (Rotary Speed, RPM)

3.3 การตั้งค่าแบบจำลองและการฝึกฝน (Model Setup and Training)

ขั้นตอนที่ 1 : การฝึกฝนการเรียนรู้แบบจำลอง (Training Machine Learning Model)

การศึกษาวิจัยนี้ ข้อมูลที่ใช้แบบตามลำดับข้อมูล (Sequential data) หรือ ความต่อเนื่องของค่าความลึก (Depth-dependent) ที่ใช้ในแบบจำลองของการดำเนินงานการขุดเจาะ ดังนั้น แบบจำลองที่ใช้ในการฝึกฝนและการทดสอบจะทำการแบ่งข้อมูลตามลำดับ (Sequential split) โดย 2 เทคนิค ที่ใช้ คือ การแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or forward-chaining split) และ การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) ทั้ง 2 เทคนิคแบ่งข้อมูลตามค่าความลึก ที่เรียกว่า การเพิ่มขนาด (Increments) ขึ้นของข้อมูล การเพิ่มขนาดขึ้นของข้อมูลเป็นตัวแทนของช่วงค่าความลึกที่ข้อมูลถูกแบ่งในแต่ละการวนซ้ำ (Iteration) ตัวอย่างของการเพิ่มขึ้น (Increments) ที่ใช้ในการวิจัย ได้แก่ 5, 10, 15, 20, 50, 100, 200 และ 500 เพื่อเลือกการทดลองที่เป็นตัวแทนการประเมินประสิทธิภาพแบบจำลองการเรียนรู้ที่ส่งผลต่อความแม่นยำ

ภาพที่ 2 แสดงการแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or forward-chaining split) และการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split) เริ่มแรก ทั้ง 2 เทคนิคมีข้อมูลฝึกฝนและการทดสอบเท่ากัน แต่ในการวนซ้ำครั้งที่ 2 การเรียนรู้แบบต่อเนื่อง (Continuous Learning or forward-chaining split) จะรวมข้อมูลใหม่เข้ากับชุดข้อมูลฝึกฝนที่มีอยู่เดิม ในขณะที่การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split) จะเลื่อนหน้าต่างของข้อมูลฝึกและข้อมูลทดสอบไปข้างหน้า โดยยังคงขนาด (Size) ของชุดข้อมูลฝึกและชุดข้อมูลทดสอบเท่ากับการเพิ่มขนาด (Increments) ขึ้นของข้อมูล



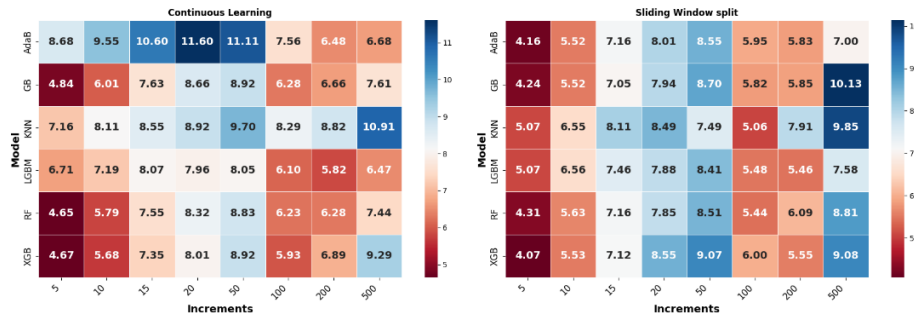
ภาพที่ 2 แสดงความแตกต่างระหว่าง การแบ่งข้อมูลแบบต่อเนื่อง (Continuous Learning or forward-chaining split) และ การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding window split)

ขั้นตอนที่ 2 : การตั้งค่าแบบจำลองและประเมินประสิทธิภาพ (Model Setup and Evaluation)

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 สำหรับการแบ่งข้อมูลตามลำดับ (Sequential split) ผู้วิจัยได้ปรับแต่งโค้ด "continuous.py" จาก GitHub [19] ให้เหมาะกับการวิจัย สำหรับการเรียนรู้แบบต่อเนื่อง (Continuous Learning) ได้ปรับแต่งเพียงจำนวนการเพิ่มขนาด (Increments) ขึ้นของข้อมูลเท่านั้น ในทางกลับกัน สำหรับการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) ได้ปรับแต่งการวนซ้ำ while เพื่อให้ชุดข้อมูลฝึกฝน (Training Set) ในรอบถัดไป มาจากชุดข้อมูลทดสอบ (Testing Set) ของรอบปัจจุบัน การประเมินแบบจำลองทั้งหมดโดยใช้ค่าความผิดพลาดสัมบูรณ์เฉลี่ย (Mean Absolute Error, MAE) ดังที่นิยามไว้ใน [20] ซึ่งต่างจากการวัดแบบอื่นๆ เช่น Mean Square Error (MSE) หรือ Root Mean Square Error (RMSE) ค่า MAE ให้การวัดความง่ายโดยตรงของค่าเฉลี่ยของข้อผิดพลาดในการทำนาย โดยไม่เน้นน้ำหนักไปที่ข้อผิดพลาดขนาดใหญ่ นอกจากนี้ยังได้ตั้งค่า Random Seed ของ Numpy และ Random State ใน scikit-learn เพื่อให้แน่ใจว่าผลลัพธ์มีความสม่ำเสมอและสามารถทำซ้ำได้ [21]

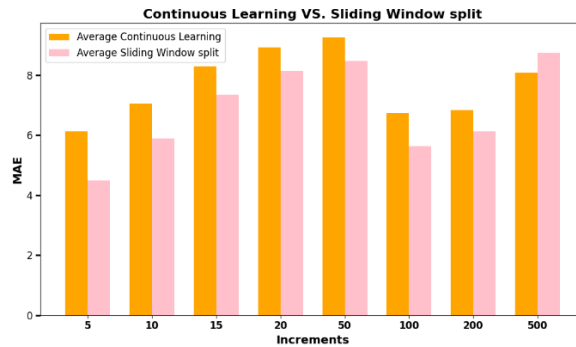
4. ผลการวิจัยและอภิปรายผลการวิจัย

ในแต่ละการวนซ้ำ (Iterations) แบบจำลองที่ทำการฝึกฝนและการทดสอบ จะใช้เทคนิคในการแบ่งข้อมูลแบบเฉพาะ (Continuous Learning and Sliding Window) กับค่าการเพิ่มขึ้น (Increments) เก็บเฉพาะค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) สำหรับชุดการทดสอบ (Testing) ค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) จากการวนซ้ำ (Iterations) ทั้งหมด นำมาคำนวณสำหรับแต่ละรอบและเทคนิคการแบ่งข้อมูลเพื่อวัดประสิทธิภาพโดยรวม จากภาพที่ 3 แสดงการเปรียบเทียบ ค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ของข้อมูลทดสอบสำหรับแบบจำลองต่างๆ กับการเพิ่มขึ้น (Increments) ของข้อมูลต่างๆ



ภาพที่ 3 แสดงค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ในแบบจำลองต่างๆ ทั้ง 8 increments

(ก) Continuous Learning (ข) Sliding Window Split

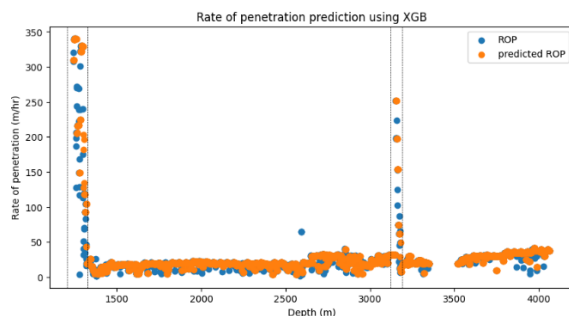


ภาพที่ 4 แสดงกราฟแท่งเปรียบเทียบค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ใน Continuous Learning และ Sliding Window Split

เพื่อดูประสิทธิภาพโดยรวมของเทคนิคการแบ่งข้อมูลแต่ละแบบ ได้มีการคำนวณค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) จากทุกแบบจำลองสำหรับแต่ละขนาดของข้อมูลที่เพิ่มขึ้น และนำเสนอในภาพที่ 4 การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window Split) มีประสิทธิภาพที่ดีกว่าในการวัดค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) สำหรับข้อมูลทดสอบ เมื่อเทียบกับการเรียนรู้แบบต่อเนื่อง (Continuous Learning) ยกเว้นการเพิ่มข้อมูลขนาด (Increments) ที่ 500 ซึ่งวิธีการเรียนรู้แบบต่อเนื่อง (Continuous Learning) แสดงค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ที่ต่ำกว่าเล็กน้อย ผลลัพธ์แสดงว่า เมื่อขนาดการเพิ่มขึ้น (Increments) ของข้อมูล แสดงค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ก็เพิ่มขึ้นเช่นกัน จนกระทั่งขนาดการเพิ่มขึ้นของข้อมูล (Increments) ถึงเท่ากับ 50 จากนั้นจะลดลงอีกครั้งที่ 100 ก่อนที่จะเพิ่มขึ้นอีกครั้ง ในการวัดประสิทธิภาพของแบบจำลองแบบผสมอย่าง Extreme Gradient Boosting (XGB), Random Forest (RF) และ Gradient Boosting (GB) แสดงค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ที่น้อยเหมือนกันกับขนาดการเพิ่มขึ้นของข้อมูล เช่น 5 และ 10 เช่นเดียวกับขนาดการเพิ่มขึ้นของข้อมูล เช่น 100 และ 200 อย่างไรก็ตาม Light Gradient Boosting Machine (LGBM) และ Adaptive Boosting (AdaBoost) แสดงค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ที่น้อยกว่าในขนาดการเพิ่มขึ้นของข้อมูล (Increments) เช่น 100, 200 และ 500

จากภาพที่ 3 แสดงว่าแบบจำลอง Extreme Gradient Boosting (XGB) ที่ใช้การแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window split) ที่ละ 5 หน่วย ได้ค่าผิดพลาดเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ที่น้อยที่สุดบนข้อมูลทดสอบ ประสิทธิภาพที่เหมาะสมของค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination, R^2) ของแบบจำลองนี้เท่ากับ 0.88 [22] จากภาพที่ 5 เปรียบเทียบค่าอัตราการเจาะ (Rate of Penetration, ROP) จริง กับค่าทำนายอัตราการเจาะโดยใช้กราฟแบบจุด

กระจาย (Scatter plot) จุดสีน้ำเงินแทนค่าอัตราการเจาะจริงจากชุดข้อมูลทั้งหมด ในขณะที่จุดสีส้มแทนค่าอัตราการเจาะที่ทำนาย เนื่องจากลักษณะของการแบ่งข้อมูลแบบหน้าต่างบานเลื่อน (Sliding Window split) ทีละ 5 หน่วย มีการฝึกฝน ทดสอบ และทำนายทั้งหมด 533 รอบ โดยเก็บเฉพาะค่าที่ทำนายเท่านั้น สังเกตว่าค่าที่ทำนายนั้นใกล้เคียงกับค่าจริง แต่มีค่าความลึกบางช่วงที่ เบี่ยงเบนไปจากการทำนาย ค่าความคลาดเคลื่อน (Residuals) ถูกคำนวณเป็นความแตกต่างระหว่างค่าจริงกับค่าที่ทำนาย เส้นประ ในภาพที่ 5 แสดงช่วงที่ค่าความคลาดเคลื่อนอยู่นอกเหนือช่วงที่ยอมรับได้ ซึ่งกำหนดไว้ที่ 3 เท่าของค่าเบี่ยงมาตรฐาน (Standard deviation) จากค่าความคลาดเคลื่อนเฉลี่ย (The average of the residuals)



ภาพที่ 5 การเปรียบเทียบระหว่างค่าจริงและค่าทำนายของอัตราการเจาะโดยแบบจำลอง Extreme Gradient Boosting (XGB)

5.สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาการใช้เทคนิคการแบ่งข้อมูลแบบตามลำดับ (Sequential data) และแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) เพื่อจำลองกระบวนการขุดเจาะและทำนายค่าอัตราการเจาะ (Rate of Penetration, ROP) โดยอิงจากข้อมูลตามความลึก ผลการวิจัยชี้ให้เห็นว่าสำหรับข้อมูลชุดนี้ เทคนิคการแบ่งหน้าต่างบานเลื่อน (Sliding Window split) ด้วยการเพิ่มขนาด (increments) ของข้อมูลเล็กๆ และแบบจำลองแบบรวมกลุ่ม (Ensemble models) ทำให้ประสิทธิภาพแบบจำลองที่ดีที่สุดในการทำนายค่าอัตราการเจาะ ในการเพิ่มขนาด (increments) ของข้อมูลขนาดเล็กน้อยนั้น ได้เปรียบเมื่อมีข้อมูลการฝึกฝนจำกัด แต่ต้องใช้เวลาวนซ้ำ (iteration) หลายครั้งเพื่อฝึกฝนแบบจำลอง เมื่อมีข้อมูลเพียงพอ การเพิ่มขนาด (increments) ขึ้นของข้อมูลสามารถลดจำนวนการวนซ้ำที่จำเป็นสำหรับชุดข้อมูลนี้ คุณสมบัติของชั้นหิน (Formation properties) เป็นปัจจัยสำคัญในการบ่งบอกการเพิ่มขนาดขึ้นของข้อมูล การเพิ่มขนาดของข้อมูลเล็กๆ หมายความว่า ข้อมูลยังคงอยู่ในชั้นหินเดียวกัน มีการศึกษาจำนวนมากที่แสดงให้เห็นว่าแบบจำลองแบบรวมกลุ่ม (Ensemble models) เช่น Extreme Gradient Boosting (XGB), Random Forest (RF), Gradient Boosting (GB) และ K-Nearest Neighbor (KNN) มักจะทำงานได้ดีในการทำนายค่าอัตราการเจาะ การศึกษานี้เน้นย้ำถึงความสำคัญของการพิจารณาข้อมูลตามความลึกและการนำไปรวมไว้ในแบบจำลองของการทำนายการขุดเจาะ หลุมสำรวจแบบเวลาจริง (Real Time) สำหรับงานในอนาคตจะเป็นการสนใจที่จะสำรวจวิธีการในการทำนายค่าความผิดพลาด (cross validation) สำหรับการปรับแต่งพารามิเตอร์ (Hyperparameter Tuning) ในแต่ละรอบการฝึกฝน

กิตติกรรมประกาศ

การจัดทำวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีจากการสนับสนุน ความรู้ ความช่วยเหลือ คำแนะนำ ตลอดจนแนวทางในการทำวิจัยและจัดทำสารนิพนธ์ของ อ.ดร.ศุภร คนธภักดิ์ อาจารย์ที่ปรึกษา และคณาจารย์ทุกท่านในหลักสูตรวิทยาการข้อมูล ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ การสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

- [1] Speer, J. W.: A Method for Determining Optimum Drilling Techniques. In: SPE Upper Gulf Coast Drilling and Production Conference, pp. SPE-1242-G. SPE, Lafayette, Louisiana (1959).
- [2] Bingham, M. G.: A New Approach to Interpreting Rock Drillability. *Oil and Gas Journal* 62, 173-179 (1964).
- [3] Hareland, G., Rampersad, P. R.: Drag-bit model including wear. In: SPE Latin America and Caribbean Petroleum Engineering Conference, pp. SPE-26957-MS. SPE, Buenos Aires, Argentina (1994).
- [4] Motahhari, H. R., Hareland, G., James, J. A.: Improved Drilling Efficiency Technique Using Integrated PDM and PDC Bit Parameters. In: *Journal of Canadian Petroleum Technology* 49(10), 45–52 (2010).
- [5] Chandrasekaran, S., Kumar, G. S.: Drilling efficiency improvement and rate of penetration optimization by machine learning and data analytics. In: *International Journal of Mathematical, Engineering and Management Sciences* 5(3), 381–394 (2020).
- [6] Hegde, C., Daigle, H., Millwater, H., Gray, K.: Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. In: *Journal of Petroleum Science and Engineering* 159, 295–306 (2017).
- [7] Soares, C., Gray, K.: Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. In: *Journal of Petroleum Science and Engineering* 172, 934-959 (2019).
- [8] Ren, Y., Lu, B., Zheng, S., Bai, K., Cheng, L., Yan, H., Wang, G.: Research on the Rate of Penetration Prediction Method Based on Stacking Ensemble Learning. In: *Geofluids* 2023, (2023).
- [9] Al-Sahlane, D.T., Allawi, R.H., Al-Mudhafar, W.J., Yao, C.: Ensemble Machine Learning for Data-Driven Predictive Analytics of Drilling Rate of Penetration (ROP) Modeling: A Case Study in a Southern Iraqi Oil Field. In: SPE Western Regional Meeting, pp. D021S004R007. SPE, Anchorage, Alaska, USA (2023).
- [10] El-Sayed, Y., Salem, A. M., El-Rammah, S.: Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms. In: *Journal of Mining and Environment* 14(3), 755–770 (2023).
- [11] Olafadehan, O. A., Ahaotu, I. D.: Comparative Analysis of Machine Learning Algorithms in Predicting Rate of Penetration during Drilling. In: *J Petro Chem Eng* 1(1), 32-47 (2023).

[12] Tunkiel, A. T., Sui, D., Wiktorski, T.: Reference dataset for rate of penetration benchmarking. In: Journal of Petroleum Science and Engineering 196, 108069 (2021).

[13] Tashman, L. J.: Out-of-sample tests of forecasting accuracy: an analysis and review. In: International Journal of Forecasting 16(4), 437–450 (2000).

[14] Equinor volve field data sharing, <https://www.equinor.com/energy/volve-data-sharing>.

[15] Understanding the WITSML format in the Volve Oilfield dataset, <https://discovervolve.com/2021/01/06/understanding-the-witsml-format-in-the-volve-oilfield-dataset/>.

[16] Interactive Well Trajectory Plot: <https://github.com/maribickpostanes/Interactive-Well-Trajectory-Plot>.

[17] Ashrafi, S.B., Anemangely, M., Sabah, M., Ameri, M.J.: Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. In: Journal of Petroleum Science and Engineering 175, 604–623 (2019).

[18] Gan, C., Cao, W. H., Wu, M., Chen, X., Hu, Y. Le, Liu, K. Z., Wang, F. W., Zhang, S. B.: Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China. In: Journal of Petroleum Science and Engineering 181, 106200 (2019).

[19] USROP: <https://github.com/AndrzejTunkiel/USROP>.

[20] Wang, W., Lu, Y.: Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. In: IOP conference series: materials science and engineering 324, 012049 (2018).

[21] Reproducibility: <https://sklearn-genetic-opt.readthedocs.io/en/stable/tutorials/reproducibility.html>.

[22] R² score, the coefficient of determination: https://scikit-learn.org/stable/modules/model_evaluation.html#r2-score-the-coefficient-of-determination.