การประยุกต์ TreeNet® Regression ในการสร้างเส้นมาตรฐานของการทดสอบการอัดแน่นดิน
Application of TreeNet® Regression for Generating a Standard Line in Soil Compaction Testing

รัตนโชติ ทองป้อง[1*], ปวีณา ชุนเกาะ[1], ธรรศ เสถียรนาม[1], ภาวิณี มากทอง[1] และ ธราพงษ์ พัฒนศักดิ์ภิญโญ[1]

Rattanachot Thongpong[1*], Paweena Chunko[1], Trust Satiennam[1],

Pawinee Markthong[1] and Tarapong Patanasakpinyo[1]

[1]หลักสูตรเทคโนโลยีวิศวกรรมโยธา, คณะเทคโนโลยีอุตสาหกรรม,

มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์, ปทุมธานี, 13180, ประเทศไทย

[1]Civil Engineering Technology Program, Faculty of Industrial Technology,

Valaya Alongkorn Rajabhat University under the Royal Patronage, Pathum Thani, 13180, Thailand

***Corresponding Author E-mail:** Rattanachot.thong@vru.ac.th

รับบทความ (Received) : November 19, 2025 /ปรับแก้ไข (Revised) : December 20, 2025 /ตอบรับบทความ (Accepted) : December 25, 2025

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการพยากรณ์ตัวแปรทางปฐพีกลศาสตร์ (Dry Density : DD) โดยอาศัยตัวแปรพยากรณ์หลักคือ Moisture content (MC) ผู้วิจัยได้ใช้แบบจำลองกลุ่มต้นไม้ (Tree-based ensemble model) กับชุดข้อมูลจำนวน 600 ตัวอย่าง และทำการปรับค่าพารามิเตอร์ที่เหมาะสมที่สุดผ่านกระบวนการตรวจสอบไขว้ 5 ส่วน (5-fold cross-validation) โดยประเมินประสิทธิภาพของแบบจำลองด้วยค่าสัมประสิทธิ์การตัดสินใจ (R-squared) และค่าความคลาดเคลื่อนเฉลี่ยสัมพัทธ์ (MAPE) ผลการวิจัยพบว่าแบบจำลองสุดท้ายที่ประกอบด้วยต้นไม้ 296 ต้น มีความสามารถในการทำนายบนชุดข้อมูลทดสอบได้อย่างน่าเชื่อถือ โดยมีค่า R-squared เท่ากับ 48.73% และ MAPE เท่ากับ 12.14% โดยไม่พบปัญหาการเรียนรู้ที่มากเกินไป (Overfitting) ที่สำคัญแบบจำลองสามารถจับความสัมพันธ์ที่มีความซับซ้อนและไม่ใช่เชิงเส้นตรง (Non-linear) ระหว่างตัวแปร MC และ DD ได้อย่างมีประสิทธิภาพ ซึ่งแสดงให้เห็นว่าอิทธิพลของ MC ต่อค่าพยากรณ์มีการเปลี่ยนแปลงไปในแต่ละช่วงค่า โดยสรุป การศึกษานี้แสดงให้เห็นถึงศักยภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์สำหรับปัญหาทางปฐพีกลศาสตร์ที่มีความสัมพันธ์ซับซ้อน และยืนยันว่า MC เป็นปัจจัยสำคัญอย่างยิ่งในการทำนายผลลัพธ์

คำสำคัญ : การเรียนรู้ของเครื่อง, การพยากรณ์, ความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง, แบบจำลองกลุ่มต้นไม้

Abtract

This research aims to develop and evaluate a machine learning framework for predicting Dry Density (DD) based on Moisture Content (MC), a fundamental relationship in geotechnical engineering. Utilizing a dataset of 600 samples, a tree-based ensemble model was developed and optimized through 5-fold cross-validation.

The model's predictive performance was quantified using the coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE). The optimized model, comprising 296 trees, achieved an R^2 of 48.73% and a MAPE of 12.14% on the test dataset. Results indicate that the model generalized well to unseen data without exhibiting signs of overfitting. Notably, the ensemble approach effectively captured the complex, non-linear, and non-monotonic relationship between MC and DD, demonstrating that the sensitivity of dry density to moisture fluctuations varies significantly across different ranges. This study concludes that machine learning techniques offer a robust alternative for modeling intricate geotechnical behaviors where traditional linear methods may fail, reaffirming moisture content as a primary predictor of soil compaction outcomes.

Keywords : Machine Learning, Prediction, Non-linear Relationship, Tree-based Ensemble Model

บทนำ

การทดสอบการอัดแน่นดินมีบทบาทสำคัญในวิศวกรรมฐานราก เพื่อสร้างความมั่นคงและความทนทานของโครงสร้างพื้นฐาน กระบวนการนี้มักเกี่ยวข้องกับการหาความสัมพันธ์ระหว่างความหนาแน่นของดินและปริมาณน้ำ เพื่อให้ได้ค่าความหนาแน่นสูงสุดของดินแห้ง (Maximum Dry Density, MDD) ที่ความชื้นเหมาะสมที่สุด (Optimum Moisture Content, OMC) กราฟการอัดแน่นที่ได้จากการทดสอบให้ข้อมูลเชิงลึกเกี่ยวกับพฤติกรรมของดิน อย่างไรก็ตาม ความคลาดเคลื่อนในการทดสอบการอัดแน่น ไม่ว่าจะเกิดจากความแตกต่างของผู้ปฏิบัติงาน การปรับเทียบเครื่องมือ หรือปัจจัยสิ่งแวดล้อม สามารถทำให้กราฟการอัดแน่นมีความแตกต่าง ส่งผลต่อความน่าเชื่อถือของการประเมินคุณสมบัติของดิน งานวิจัย เช่น [1-3] ได้เน้นถึงความหลากหลายขององค์ประกอบดิน สภาพแวดล้อม และความคลาดเคลื่อนจากผู้ปฏิบัติงาน ที่ส่งผลต่อผลลัพธ์การอัดแน่น จึงเน้นย้ำความจำเป็นในการพัฒนาวิธีการทดสอบที่มีความเสถียรมากขึ้น

เพื่อแก้ไขความไม่สอดคล้องเหล่านี้ วิธีการในปัจจุบันมักมุ่งเน้นไปที่การสร้าง เส้นมาตรฐาน เพื่อแทนชุดกราฟการอัดแน่นให้เป็นโปรไฟล์เดียวที่สอดคล้องกัน วิธีนี้มีความสำคัญต่อการทำให้ผลลัพธ์การอัดแน่นเป็นมาตรฐานระหว่างการทดสอบและผู้ปฏิบัติงานที่แตกต่างกัน วิธีที่ใช้บ่อยในการสร้างเส้นมาตรฐาน ได้แก่ การหาค่าเฉลี่ยของกราฟการอัดแน่นและการสร้างสมการพหุนามเพื่อแทนข้อมูล นอกจากนี้วิธีการหาค่าเฉลี่ยตามงานวิจัย [4] จะสร้างเส้นฐานโดยคำนวณค่าเฉลี่ยของกราฟการอัดแน่นหลายครั้ง ขณะที่การสร้างสมการพหุนามตามงานวิจัย [5-6] พยายามแทนแนวโน้มโดยรวมของข้อมูลการอัดแน่นผ่านแบบจำลองทางคณิตศาสตร์ แม้ว่าวิธีเหล่านี้จะให้ภาพรวมที่เรียบง่าย แต่ก็อาจไม่สามารถสะท้อนความหลากหลายและรายละเอียดของข้อมูลการอัดแน่นได้อย่างครบถ้วน งานวิจัยล่าสุด เช่น [7-9] ได้ศึกษาวิธีทางสถิติและเชิงคำนวณในการสร้างเส้นมาตรฐาน แต่ยังมีข้อจำกัดในการจัดการกับชุดข้อมูลที่หลากหลายอย่างมีประสิทธิภาพ

ในปัจจุบัน การเรียนรู้ของเครื่อง (Machine Learning: ML) ได้เข้ามามีบทบาทสำคัญในการเพิ่มขีดความสามารถของการวิเคราะห์ทางวิศวกรรมโยธาและปฐพีกลศาสตร์ โดยเฉพาะการนำมาใช้จัดการกับข้อมูลที่มีความซับซ้อนและมีความสัมพันธ์เชิงไม่เชิงเส้น (Non-linear Relationship) ซึ่งวิธีการทางสถิติแบบดั้งเดิมอาจไม่สามารถอธิบายพฤติกรรมได้อย่างครอบคลุม

ในบรรดาเทคนิคต่างๆ ของ ML นั้น วิธีการ แบบรวมกลุ่ม (Ensemble Learning) ได้รับการยอมรับอย่างกว้างขวางว่าให้ความแม่นยำสูงกว่าการใช้โมเดลเดี่ยว เนื่องจากเป็นการรวมจุดแข็งของหลายแบบจำลองเข้าด้วยกันเพื่อลดความคลาดเคลื่อน โดยเฉพาะอย่างยิ่งในงานวิจัยด้านวิศวกรรมฐานราก [10-12] ที่พบว่าการสร้างแบบจำลองพฤติกรรมของดินมีความท้าทายจากความผันแปรของข้อมูลสูง

ด้วยเหตุนี้ TreeNet® Regression ซึ่งเป็นอัลกอริทึมในกลุ่ม Gradient Boosting จึงเป็นเครื่องมือที่มีความเหมาะสมและมีแนวโน้มที่ดีมากสำหรับการนำมาใช้สร้างเส้นมาตรฐานของกราฟการอัดแน่นดิน (Soil Compaction Curve) เนื่องจาก

ความสามารถที่โดดเด่นในการสร้างแบบจำลองจากชุดข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ โดยการฝึกฝนโมเดล TreeNet® ด้วยข้อมูลจากการทดสอบการอัดแน่น จะช่วยให้สามารถสร้างเส้นมาตรฐานที่สะท้อนพฤติกรรมที่แท้จริงของดินได้แม่นยำกว่าวิธีพหุนาม (Polynomial) หรือวิธีการหาค่าเฉลี่ยแบบดั้งเดิม อีกทั้งยังช่วยลดความคลาดเคลื่อนและรองรับความหลากหลายของข้อมูลจากการทดสอบซ้ำหลายครั้งได้อย่างน่าเชื่อถือ วัตถุประสงค์ของงานวิจัยนี้คือการประยุกต์ใช้ TreeNet® Regression เพื่อสร้างเส้นมาตรฐานของการทดสอบการอัดแน่นดิน และประเมินประสิทธิภาพของโมเดลโดยใช้ตัวชี้วัดทางสถิติพื้นฐาน เช่น R^2, RMSE และ MAE วิธีการนี้มุ่งเน้นการเพิ่มความแม่นยำ ความเชื่อถือได้ และการสร้างมาตรฐานในการประเมินการอัดแน่นดิน เพื่อสนับสนุนการปฏิบัติทางวิศวกรรมฐานรากที่มีประสิทธิผลยิ่งขึ้น

1. วัตถุประสงค์การวิจัย

1.1 เพื่อศึกษาความเป็นไปได้ ในการสร้างแบบจำลองในการสร้างเส้นมาตรฐานของกราฟการอัดแน่นดิน จากข้อมูลการทดสอบหลายชุด

1.2 เพื่อประเมินความแม่นยำและความน่าเชื่อถือของเส้นมาตรฐานที่สร้างขึ้น โดยใช้ตัวชี้วัดทางสถิติพื้นฐาน เช่น R^2, RMSE และ MAE

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

การทดสอบการอัดแน่นดินเป็นกระบวนการสำคัญในวิศวกรรมฐานราก โดยความน่าเชื่อถือของผลการทดสอบอาจได้รับผลกระทบจากหลายปัจจัย Zhang et al. (2021) [1] ได้ระบุถึงความแตกต่างขององค์ประกอบดิน โดยแสดงให้เห็นว่าความแปรผันของแร่ธาตุและขนาดอนุภาคส่งผลต่อผลลัพธ์ของการทดสอบอย่างมีนัยสำคัญ ในทำนองเดียวกัน Mahajan et al. (2020) [2] ได้ศึกษาผลกระทบของสภาพแวดล้อม เช่น อุณหภูมิและความชื้น ต่อการวัดค่าการอัดแน่นดิน และพบว่าการเปลี่ยนแปลงเพียงเล็กน้อยของสภาวะแวดล้อมสามารถทำให้ข้อมูลคลาดเคลื่อนได้ งานวิจัยเหล่านี้ชี้ให้เห็นถึงความสำคัญของการพิจารณาความหลากหลายของดินและปัจจัยภายนอกในการวิเคราะห์ผลการทดสอบ

นอกเหนือจากความแตกต่างตามธรรมชาติ ปัจจัยจากผู้ปฏิบัติงานยังมีผลต่อความแม่นยำของผลลัพธ์ โดย Miller et al. (2019) [3] ได้ศึกษาความคลาดเคลื่อนที่เกิดจากผู้ปฏิบัติงาน และเน้นย้ำว่าความแตกต่างในการดำเนินขั้นตอนการทดสอบรวมถึงการตัดสินใจเฉพาะหน้า สามารถสร้างความไม่สอดคล้องระหว่างชุดข้อมูลได้ งานวิจัยดังกล่าวนำเสนอความจำเป็นในการใช้มาตรฐานการทดสอบเพื่อลดความคลาดเคลื่อน และสะท้อนให้เห็นถึงข้อจำกัดของวิธีการดั้งเดิมที่อาศัยการปฏิบัติงานด้วยมือร่วมกับการหาค่าเฉลี่ยทางสถิติแบบพื้นฐาน โดยสรุปแล้ว แหล่งที่มาของความแปรปรวนอันประกอบด้วย คุณสมบัติของดิน สภาพแวดล้อม และพฤติกรรมของผู้ปฏิบัติงาน ล้วนเป็นปัจจัยที่ทำให้การตีความและการเปรียบเทียบกราฟการอัดแน่นดินทำได้ยาก

เพื่อแก้ไขปัญหาดังกล่าว งานวิจัยในปัจจุบันจึงมุ่งเน้นไปที่การสร้าง "เส้นมาตรฐาน" (Baseline Curve) เพื่อรวมชุดกราฟการอัดแน่นหลายชุดให้เป็นโปรไฟล์เดียว โดย Singh et al. (2021) [4] ได้ศึกษาวิธีการทางสถิติ เช่น เทคนิคการหาค่าเฉลี่ยและการสร้างแบบจำลองเชิงถดถอย เพื่อสร้างกราฟตัวแทนที่สะท้อนความแตกต่างระหว่างการทดสอบ ต่อมา Kim et al. (2020) [5] ได้พัฒนาวิธีการเชิงคำนวณเพื่อบริหารจัดการความแตกต่างของกราฟ โดยแสดงให้เห็นว่าวิธีการเชิงอัลกอริทึมสามารถตรวจจับแนวโน้มเชิงไม่เชิงเส้นและความแตกต่างเพียงเล็กน้อยระหว่างการทดสอบได้ดีกว่าการหาค่าเฉลี่ยเพียงอย่างเดียว งานวิจัยเหล่านี้สะท้อนถึงแนวโน้มการใช้เครื่องมือเชิงคำนวณขั้นสูงเพื่อปรับปรุงมาตรฐานและความน่าเชื่อถือของข้อมูล

อย่างไรก็ตาม การรวมกราฟการอัดแน่นจากชุดข้อมูลที่หลากหลายยังคงเป็นความท้าทาย Arshad et al. (2019) [6] ได้ระบุถึงข้อจำกัดของวิธีการในปัจจุบันและความยากลำบากในการจัดการกับดินประเภทต่างๆ รวมถึงมาตรฐานการทดสอบที่แตกต่างกัน งานวิจัยนี้เน้นว่าแม้การสร้างเส้นมาตรฐานจะช่วยเพิ่มความสอดคล้องของข้อมูลได้ แต่จำเป็นต้องเลือกเทคนิคการ

สร้างแบบจำลองที่เหมาะสมเพื่อให้ได้ผลลัพธ์ที่แม่นยำและสามารถประยุกต์ใช้ได้ทั่วไป ซึ่งถือเป็นรากฐานสำคัญในการพัฒนาแบบจำลองพยากรณ์ขั้นสูง เช่น TreeNet® Regression เพื่อสร้างเส้นมาตรฐานที่มีความแข็งแรง (Robust) และรองรับความหลากหลายของสภาพดินได้อย่างมีประสิทธิภาพ

วิธีดำเนินการวิจัย

1. ข้อมูลดินและขั้นตอนการทดลอง

ดินที่ใช้ในการศึกษานี้จัดอยู่ในประเภท GW (Well-Graded Gravel) ตามระบบการจำแนกดินเอกภาพ (USCS) และจัดเป็น A-2-5 (Clayey Gravel) ตามระบบของ AASHTO โดยลักษณะทางกายภาพเป็นกรวดคละชั้นดี มีปริมาณตะกอนดินเหนียวปนเล็กน้อย ซึ่งยืนยันได้ว่าเป็นวัสดุที่มีคุณสมบัติเหมาะสมสำหรับใช้เป็นชั้นรองพื้นทางตามมาตรฐานงานทางหลวง สมบัติพื้นฐานของดินประกอบด้วย ค่าขีดจำกัดเหลว (Liquid Limit) เท่ากับ 49.16 ดัชนีความเหนียว (Plasticity Index) เท่ากับ 3.89 และความถ่วงจำเพาะ (Specific Gravity) อยู่ในช่วง 2.65–2.70

สำหรับการเตรียมตัวอย่าง ดินถูกร่อนผ่านตะแกรงมาตรฐานหมายเลข 4 (4.75 มม.) และปรับเพิ่มความชื้นตามลำดับก่อนการอัดแน่น การทดลองใช้วิธีการอัดแน่นแบบสูงกว่ามาตรฐาน (Modified Proctor Test) ตามข้อกำหนด ASTM D1557 โดยใช้แม่พิมพ์ทรงกระบอก แบ่งดินออกเป็น 5 ชั้น และอัดด้วยลูกตุ้มน้ำหนัก 4.54 กก. ตกจากความสูง 457 มม. เพื่อหาค่าความหนาแน่นแห้งสูงสุด (Maximum Dry Density: MDD) และปริมาณความชื้นที่เหมาะสม (Optimum Moisture Content: OMC)

การวิเคราะห์ใช้ฐานข้อมูลจากการทดสอบทั้งหมด 120 ชุด โดยแต่ละชุดมีการทดสอบย่อย 5 ระดับความชื้น (2%, 4%, 6%, 8% และ 10%) รวมเป็นข้อมูลทั้งหมด 600 ตัวอย่าง ดำเนินการโดยกลุ่มผู้ปฏิบัติงาน 3 กลุ่ม เพื่อสร้างความหลากหลายของข้อมูลและสะท้อนความแปรปรวนที่เกิดขึ้นจริงในภาคสนาม ข้อมูลจำนวน 600 ตัวอย่างนี้ ถือว่ามีความเพียงพอสำหรับการพัฒนาแบบจำลองการเรียนรู้ของเครื่องในงานเฉพาะด้านปฐพีวิศวกรรม เนื่องจากเป็นข้อมูลที่ได้จากการควบคุมสภาพแวดล้อมในห้องปฏิบัติการ (Experimental Data) ซึ่งมีความหนาแน่นของสารสนเทศสูงกว่าข้อมูลทั่วไป ช่วยให้โมเดลสามารถเรียนรู้รูปแบบความสัมพันธ์ได้โดยไม่เกิดปัญหาการขาดแคลนข้อมูล

ตารางที่ 1 ข้อมูลทางสถิติจากการทดลอง

รายละเอียด		มากที่สุด	น้อยที่สุด	ลำดับกลุ่มข้อมูลหรือช่วงของการทดสอบ				ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
				Q1	Q2	Q3	Q4		
ความหนาแน่นแห้งของดิน (Dry Density : DD)	**2%**	1.67	0.69	0.80	0.97	1.03	1.67	0.95	0.40
	4%	1.88	1.21	1.38	1.49	1.60	1.88	1.51	0.25
	6%	1.96	1.41	1.55	1.67	1.80	1.96	1.67	0.20
	8%	1.64	1.14	1.26	1.32	1.43	1.64	1.35	0.19
	10%	1.58	0.96	1.04	1.12	1.19	1.58	1.14	0.25
ปริมาณความชื้นในดิน (Moisture Content : MC)	**2%**	18.15	4.45	5.01	8.87	13.12	18.15	9.19	5.68
	4%	25.63	9.93	11.79	13.07	16.76	25.63	14.39	6.43
	6%	27.09	12.50	15.66	20.61	22.96	27.09	19.60	5.48
	8%	34.77	14.51	20.01	25.95	29.21	34.77	24.84	7.45

10%	48.07	15.54	22.96	36.65	40.08	48.07	32.78	12.24
ความหนาแน่นแห้งสูงสุด (Maximum Dry Density : MDD)	1.9552	1.96	1.41	1.55	1.68	1.81	1.96	1.68
ปริมาณความชื้นที่เหมาะสมในการอัดแน่น (Optimum Moisture Content : OMC)	27.0851	27.09	9.93	13.80	16.98	22.07	27.09	17.89
ค่าพลังงาน	1478.2900	1478.29	782.65	992.41	1077.75	1182.86	1478.29	1092.70

2. การวิเคราะห์โดยใช้ TreeNet® Regression

TreeNet® Regression เป็นเทคนิคการสร้างแบบจำลองเชิงพยากรณ์ที่พัฒนาบนพื้นฐานของอัลกอริทึมต้นไม้ตัดสินใจแบบเพิ่มระดับความชัน (Gradient Boosting Decision Tree: GBDT)

ความเหมาะสมของอัลกอริทึมกับงานวิจัย: สาเหตุที่เลือกใช้ TreeNet® สำหรับการสร้างเส้นมาตรฐานการอัดแน่นดิน เนื่องจากพฤติกรรมความสัมพันธ์ระหว่างความชื้นและความหนาแน่นของดินมีความซับซ้อนและเป็นเชิงไม่เชิงเส้น (Non-linear) อีกทั้งข้อมูลจากการทดสอบมักมีความคลาดเคลื่อน (Noise) จากปัจจัยสภาพแวดล้อมและผู้ปฏิบัติงาน TreeNet® มีจุดเด่นในการสร้างชุดต้นไม้ตัดสินใจหลายต้นแบบต่อเนื่อง โดยต้นไม้ลำดับถัดไปจะถูกสร้างขึ้นเพื่อประมาณค่าความคลาดเคลื่อนสะสม (Residuals) จากต้นไม้ลำดับก่อนหน้า ทำให้สามารถขจัดข้อผิดพลาดและจับรูปแบบข้อมูลที่ละเอียดอ่อนได้ดีกว่าโมเดลเชิงเส้นทั่วไป

กระบวนการของอัลกอริทึมดังนี้

1) การกำหนดค่าเริ่มต้น: กำหนดค่าพยากรณ์แรกเริ่มจากค่าเฉลี่ยของความหนาแน่นดินในชุดข้อมูล
2) การคำนวณค่าส่วนต่าง: คำนวณหาค่าความคลาดเคลื่อนระหว่างค่าที่วัดจริงกับค่าที่พยากรณ์ได้
3) การสร้างต้นไม้ลำดับถัดไป: สร้างต้นไม้ตัดสินใจขนาดเล็กเพื่อเรียนรู้และพยากรณ์ค่าความคลาดเคลื่อนนั้น
4) การปรับปรุงโมเดล: รวมผลลัพธ์ของต้นไม้ใหม่เข้ากับโมเดลเดิม โดยใช้ อัตราการเรียนรู้ (Learning Rate) เพื่อควบคุมความเร็วในการปรับปรุง ป้องกันไม่ให้โมเดลยึดติดกับข้อมูลชุดฝึกสอนมากเกินไป (Overfitting)
5) การทำซ้ำ: ดำเนินการวนซ้ำจนกระทั่งค่าความคลาดเคลื่อนลดลงจนอยู่ในระดับที่ยอมรับได้

3. ตัวชี้วัดประสิทธิภาพ

ในการประเมินความแม่นยำของเส้นมาตรฐานที่สร้างขึ้นจาก TreeNet® Regression งานวิจัยนี้ได้ใช้ตัวชี้วัดทางสถิติเพื่อยืนยันประสิทธิภาพของแบบจำลอง ได้แก่

1) สัมประสิทธิ์การตัดสินใจ (R^2): เพื่อวัดระดับความสามารถในการอธิบายความแปรผันของข้อมูล
2) รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE): เพื่อวัดค่าความคลาดเคลื่อนในหน่วยเดียวกับข้อมูลจริง
3) ค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE): เพื่อดูขนาดของความผิดพลาดเฉลี่ยจากการพยากรณ์

ผลการวิจัย

จากการศึกษาครั้งนี้ ผู้วิจัยได้สร้างแบบจำลองการพยากรณ์โดยกำหนดค่าพารามิเตอร์ต่างๆ ของแบบจำลอง (ดังแสดงในภาพ) โดยใช้ฟังก์ชันการสูญเสีย (Loss function) เป็น Squared error และเลือกจำนวนต้นไม้ที่เหมาะสมที่สุดโดยพิจารณาจากค่า Maximum R-squared ผ่านกระบวนการตรวจสอบความถูกต้องของแบบจำลองด้วยวิธี 5-fold cross-validation ทั้งนี้ ได้กำหนด

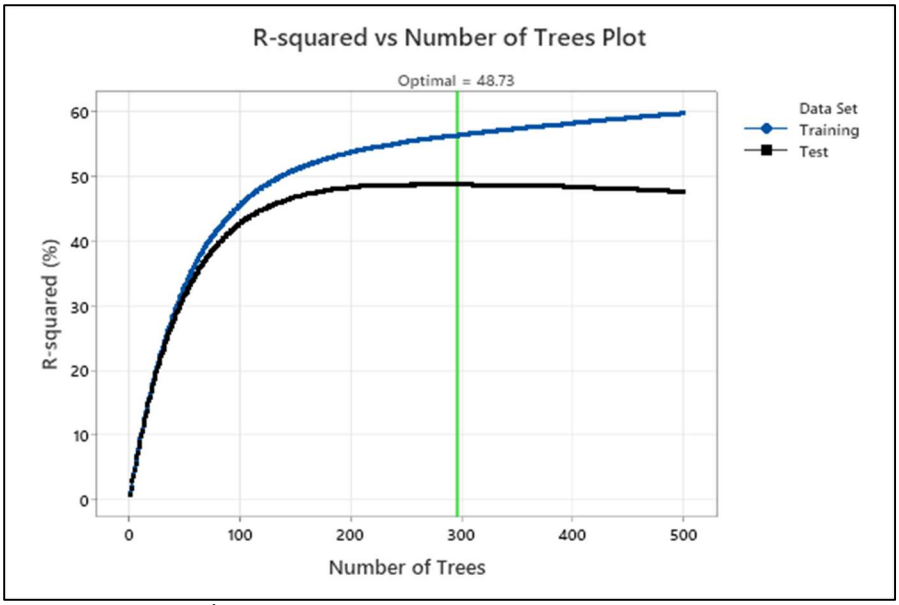
ค่าพารามิเตอร์ที่สำคัญ ได้แก่ Learning rate เท่ากับ 0.01, Subsample fraction เท่ากับ 0.5, Maximum terminal nodes per tree เท่ากับ 6 และ Minimum terminal node size เท่ากับ 3 โดยแบบจำลองดังกล่าวสร้างขึ้นจากข้อมูลที่ใช้ในการวิเคราะห์จำนวนทั้งสิ้น 600 แถว

จากผลการวิเคราะห์สถิติเชิงพรรณนาของตัวแปรตาม (Response Information) พบว่าข้อมูลมีค่าเฉลี่ย (Mean) อยู่ที่ 1.325 และมีส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เท่ากับ 0.293 ซึ่งบ่งชี้ถึงการกระจายตัวของข้อมูลรอบค่าเฉลี่ย โดยมีค่ามัธยฐาน (Median) ซึ่งเป็นค่ากลางของข้อมูลอยู่ที่ 1.319 และมีพิสัยของข้อมูลตั้งแต่ค่าต่ำสุด (Minimum) ที่ 0.695 ไปจนถึงค่าสูงสุด (Maximum) ที่ 1.955

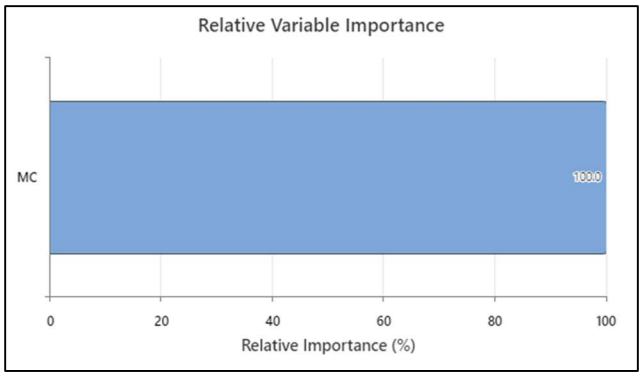
จากรูปที่ 1 แสดงผลการวิเคราะห์เพื่อหาจำนวนต้นไม้ (Number of Trees) ที่เหมาะสมที่สุดสำหรับแบบจำลอง พบว่าค่าสัมประสิทธิ์การตัดสินใจ (R-squared) ของชุดข้อมูลทดสอบ (Test set) เพิ่มขึ้นอย่างรวดเร็วจนมีค่าสูงสุดที่ 48.73% เมื่อจำนวนต้นไม้มีค่าประมาณ 280-300 ต้น หลังจากจุดดังกล่าว การเพิ่มจำนวนต้นไม้เข้าไปอีกไม่ได้ช่วยให้ประสิทธิภาพของแบบจำลองดีขึ้นอย่างมีนัยสำคัญและมีแนวโน้มคงที่ ในขณะที่ค่า R-squared ของชุดข้อมูลฝึกฝน (Training set) ยังคงเพิ่มขึ้นอย่างต่อเนื่อง ดังนั้น ผู้วิจัยจึงเลือกจำนวนต้นไม้ที่เหมาะสมที่สุด (Optimal) สำหรับการสร้างแบบจำลองนี้ที่ประมาณ 300 ต้น ซึ่งเป็นจุดที่ให้ประสิทธิภาพสูงสุดในการพยากรณ์ข้อมูลใหม่

จากสรุปผลการสร้างแบบจำลอง (Model Summary) พบว่าแบบจำลองที่เหมาะสมที่สุดประกอบด้วยจำนวนต้นไม้ 296 ต้น (Optimal number of trees) และใช้ตัวพยากรณ์ที่สำคัญ 1 ตัว เมื่อนำแบบจำลองไปทดสอบกับชุดข้อมูลทดสอบ (Test set) พบว่าแบบจำลองสามารถอธิบายความผันแปรของข้อมูลได้ร้อยละ 48.73 (R-squared = 48.73%) และมีความคลาดเคลื่อนในการทำนายโดยเฉลี่ยอยู่ที่ร้อยละ 12.14 (MAPE = 0.1214) โดยมีค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) เท่ากับ 0.2094 ซึ่งเป็นค่าสถิติที่ใกล้เคียงกับผลลัพธ์ในชุดข้อมูลฝึกฝน (Training set) ชี้ให้เห็นว่าแบบจำลองที่พัฒนาขึ้นนี้มีประสิทธิภาพและมีความสามารถในการนำไปใช้พยากรณ์กับข้อมูลชุดใหม่ได้ดี

จากรูปที่ 2 แสดงผลการวิเคราะห์ความสำคัญสัมพัทธ์ของตัวแปร (Relative Variable Importance) ที่มีอิทธิพลต่อแบบจำลอง พบว่าตัวแปร MC เป็นตัวพยากรณ์เพียงตัวเดียวที่มีความสำคัญสูงสุด โดยมีค่าความสำคัญสัมพัทธ์คิดเป็น 100% ผลลัพธ์นี้บ่งชี้ว่าความสามารถในการพยากรณ์ทั้งหมดของแบบจำลองที่พัฒนาขึ้นนี้ขึ้นอยู่กับอิทธิพลของตัวแปร MC เพียงตัวเดียว



รูปที่ 1 สรุปผลการสร้างแบบจำลอง (Model Summary)



รูปที่ 2 แสดงผลการวิเคราะห์ความสำคัญสัมพัทธ์ของตัวแปร (Relative Variable Importance)

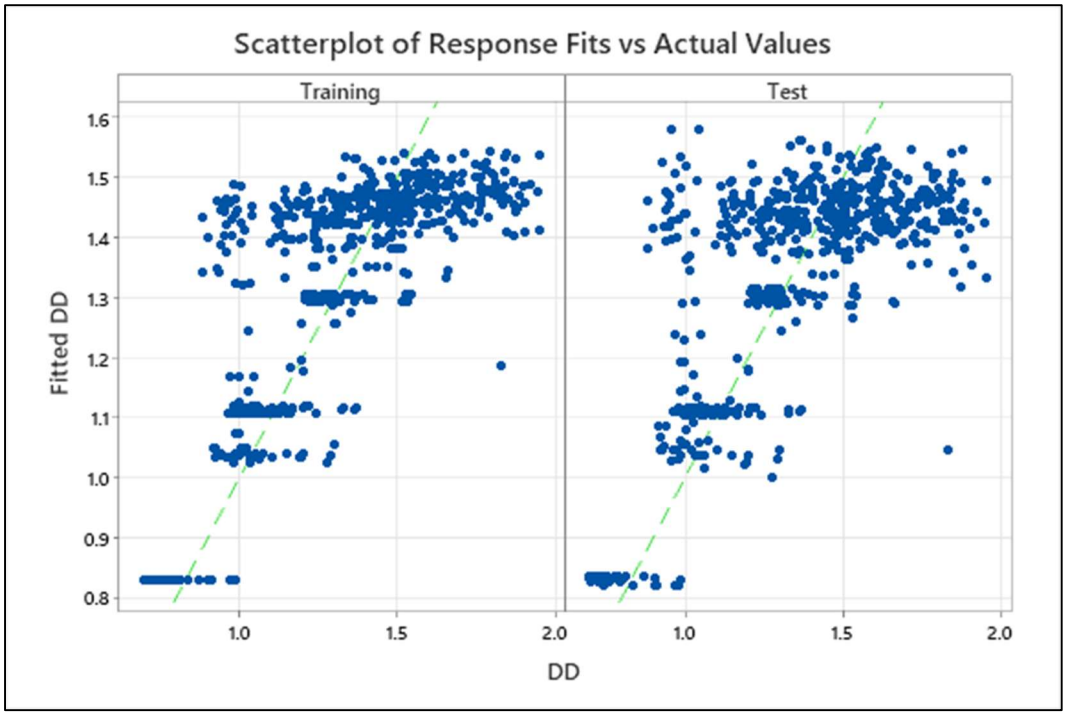
จากรูปที่ 3 แสดงผลการประเมินความแม่นยำของแบบจำลองด้วยแผนภาพการกระจาย (Scatterplot) ซึ่งเปรียบเทียบระหว่างค่าที่ทำนายได้ (Fitted DD) กับค่าจริง (DD) ดังแสดงในภาพ พบว่าจุดข้อมูลทั้งในชุดข้อมูลฝึกฝน (Training) และชุดข้อมูลทดสอบ (Test) มีการกระจายตัวเกาะกลุ่มตามแนวเส้นทแยงมุม ซึ่งเป็นแนวที่ค่าทำนายและค่าจริงมีค่าเท่ากัน อย่างไรก็ตาม ยังพบการกระจายตัวของค่าความคลาดเคลื่อนอยู่พอสมควร โดยมีลักษณะการเกาะกลุ่มของข้อมูลเป็นแถบในแนวนอน ทั้งนี้ รูปแบบการกระจายของข้อมูลในชุดข้อมูลทดสอบมีความคล้ายคลึงกับชุดข้อมูลฝึกฝนอย่างมาก ซึ่งบ่งชี้ว่าแบบจำลองที่สร้างขึ้นไม่มีปัญหาการเรียนรู้ที่มากเกินไป (Overfitting) และมีความสามารถในการสรุปผลเพื่อนำไปใช้ทำนายข้อมูลชุดใหม่ได้ดีในระดับหนึ่ง

จากรูปที่ 4 แสดงแผนภาพการพึ่งพาบางส่วน (Partial Dependence Plot) ซึ่งแสดงอิทธิพลของตัวแปร MC ต่อค่าเฉลี่ยของค่าพยากรณ์ (Fitted DD) พบว่าความสัมพันธ์ระหว่างตัวแปรทั้งสองมีลักษณะที่ไม่ใช่เชิงเส้นตรง (Non-linear) อย่างชัดเจน โดยในช่วงที่ค่า MC เพิ่มขึ้นจากประมาณ 5 ถึง 10 ส่งผลให้ค่าพยากรณ์เพิ่มสูงขึ้นอย่างรวดเร็วจนถึงระดับสูงสุด หลังจากนั้นค่าพยากรณ์จะค่อนข้างคงที่เมื่อ MC อยู่ในช่วง 10 ถึง 25 ก่อนจะลดลงอย่างรวดเร็วและกลับมามีค่าคงที่อีกครั้งที่ประมาณ 1.1 เมื่อค่า MC มากกว่า 32 ขึ้นไป ซึ่งแสดงให้เห็นว่าอิทธิพลของตัวแปร MC ต่อผลลัพธ์ของแบบจำลองนั้นมีความซับซ้อนและเปลี่ยนแปลงไปตามแต่ละช่วงค่าของตัวมันเอง

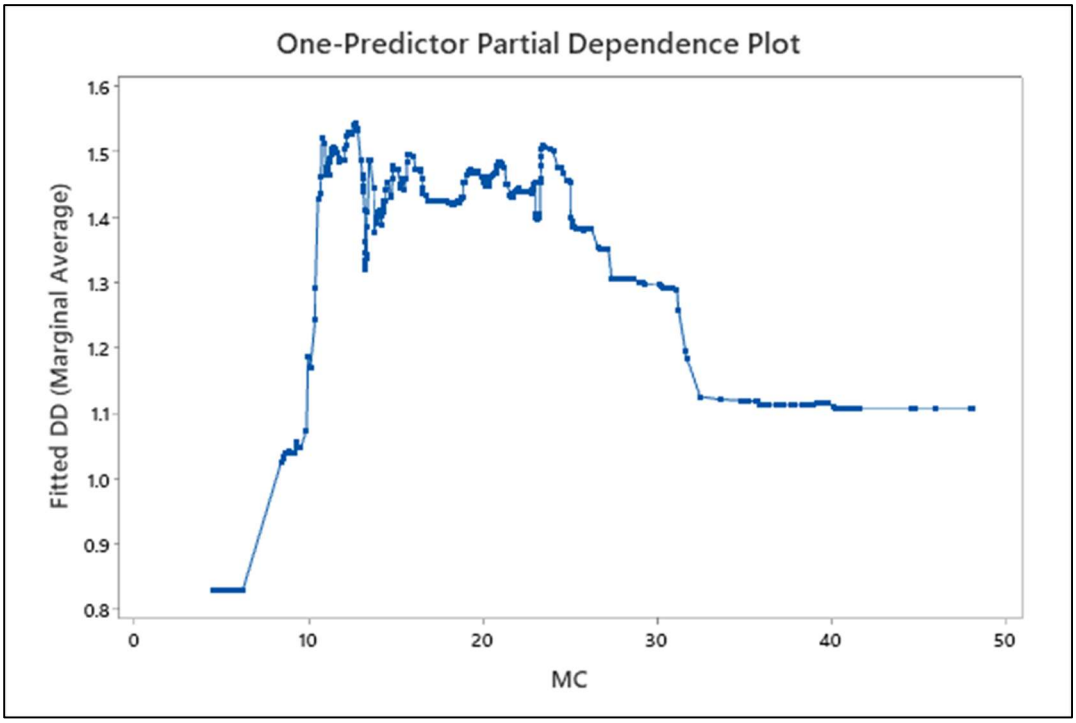
อภิปรายผล

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อทำนายตัวแปรตาม (DD) จากตัวแปรอิสระ (MC) ซึ่งผลการวิเคราะห์ให้ข้อมูลเชิงลึกที่สำคัญหลายประการเมื่อเปรียบเทียบกับองค์ความรู้ในงานวิจัยที่เกี่ยวข้อง

แบบจำลองที่พัฒนาขึ้นมีความสามารถในการอธิบายความผันแปรของข้อมูลในชุดข้อมูลทดสอบได้ร้อยละ 48.73 (R-squared) และมีค่าความคลาดเคลื่อนในการพยากรณ์เฉลี่ยร้อยละ 12.14 (MAPE) ซึ่งบ่งชี้ถึงประสิทธิภาพในระดับปานกลางที่สามารถนำไปใช้ประโยชน์ได้จริงในทางปฏิบัติ จุดแข็งที่สำคัญของแบบจำลองนี้คือ ความสามารถในการสรุปผลไปยังข้อมูลใหม่ (Generalization) ดังจะเห็นได้จากค่าสถิติและลักษณะการกระจายตัวในแผนภาพ Scatterplot (รูปที่ 4) ของชุดข้อมูลทดสอบ (Test set) ที่มีความคล้ายคลึงกับชุดข้อมูลฝึกฝน (Training set) อย่างมาก สิ่งนี้ชี้ให้เห็นว่าแบบจำลองไม่ได้เกิดภาวะการเรียนรู้ที่มากเกินไป (Overfitting) ซึ่งเป็นปัญหาที่พบได้บ่อยในแบบจำลองที่ซับซ้อน และสะท้อนให้เห็นว่าแบบจำลองได้เรียนรู้รูปแบบที่แท้จริงของข้อมูล ไม่ใช่เพียงการท่องจำข้อมูลฝึกฝนเท่านั้น



รูปที่ 3 แผนภาพการกระจาย (Scatterplot) เปรียบเทียบระหว่างค่าที่ทำนายได้ (Fitted DD) กับค่าจริง (DD)



รูปที่ 4 แผนภาพการพึ่งพาบางส่วนที่แสดงอิทธิพลของตัวแปร MC ต่อค่าเฉลี่ยของค่าพยากรณ์

ผลการวิเคราะห์ความสำคัญของตัวแปร (Relative Variable Importance) ยืนยันว่า MC เป็นตัวแปรพยากรณ์เพียงตัวเดียวที่มีนัยสำคัญ (100%) ในการทำนายค่า DD การค้นพบนี้เน้นย้ำถึงความจำเป็นในการวัดค่า MC ให้มีความแม่นยำสูงสุด เนื่องจากความคลาดเคลื่อนเพียงเล็กน้อยของตัวแปรนี้จะส่งผลกระทบโดยตรงต่อประสิทธิภาพของแบบจำลอง ซึ่งสอดคล้องกับงานวิจัยของ Zhang, X., et al. (2021) [1], Mahajan, R., et al. (2020) [2] และ Miller, T., et al. (2019) [3] ที่ชี้ให้เห็นว่าความแปรปรวนของผลการทดสอบการบดอัดดินมักเกิดจากปัจจัยต่างๆ เช่น ส่วนประกอบของดิน สภาพแวดล้อม หรือแม้กระทั่งผู้ปฏิบัติงาน ซึ่งตัวแปร MC ในงานวิจัยนี้อาจเป็นตัวแทนของปัจจัยที่ซับซ้อนเหล่านี้

ประเด็นที่น่าสนใจที่สุดคือ ความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง (Non-linear Relationship) ระหว่าง MC และ DD ดังที่แสดงในแผนภาพการพึ่งพาบางส่วน (Partial Dependence Plot) รูปที่ 5 ซึ่งแสดงให้เห็นว่าอิทธิพลของ MC ต่อค่า DD นั้นเปลี่ยนแปลงไปตามช่วงค่าต่างๆ อย่างซับซ้อน ทั้งช่วงที่ส่งผลให้ค่าพยากรณ์เพิ่มขึ้นอย่างรวดเร็ว, ช่วงที่คงที่ และช่วงที่ลดลง ซึ่งความสัมพันธ์ลักษณะนี้ยากที่จะอธิบายได้ด้วยแบบจำลองทางสถิติแบบดั้งเดิม เช่น การถดถอยพหุนาม (Polynomial Regression) ที่กล่าวถึงในงานวิจัยของ Chen, W., et al. (2020) [5] และ Gupta, N., et al. (2019) [6] หรือวิธีที่อิงค่าเฉลี่ยของ Zhou, X., et al. (2021) [4] การใช้เทคนิค Machine Learning ในงานวิจัยนี้จึงมีความเหมาะสมอย่างยิ่ง เพราะสามารถจับรูปแบบที่ซับซ้อนเหล่านี้ได้ดีกว่า

การประยุกต์ใช้เทคนิค Machine Learning ในงานวิจัยนี้สอดคล้องกับแนวโน้มสมัยใหม่ในสาขาวิศวกรรมปฐพีที่เริ่มนำแบบจำลองขั้นสูงมาใช้ในการแก้ปัญหา ดังเช่นงานวิจัยของ Li, J., et al. (2022) [10], Alam, M., et al. (2021) [11] และ Kumar, R., et al. (2020) [12] ที่ประสบความสำเร็จในการใช้โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) ในการทำนายคุณสมบัติการบดอัดของดิน การศึกษานี้ได้ช่วยเสริมองค์ความรู้ในด้านนี้โดยแสดงให้เห็นถึงศักยภาพของแบบจำลองกลุ่มต้นไม้ (Tree-based ensemble model) ซึ่งเป็นอีกทางเลือกหนึ่งที่มีประสิทธิภาพและให้ผลลัพธ์ที่สามารถตีความความสัมพันธ์ของตัวแปรได้ง่ายผ่านเครื่องมืออย่าง Partial Dependence Plot แม้ว่าแบบจำลองจะแสดงให้เห็นถึงประสิทธิภาพที่น่าพอใจ แต่ยังมีข้อจำกัดบางประการ ประการแรกคือ แบบจำลองอาศัยตัวแปร MC เพียงตัวเดียว ซึ่งทำให้ค่า R-squared อยู่ในระดับปานกลาง (48.73%) และหมายความว่ายังมีความผันแปรอีกกว่า 50% ที่แบบจำลองยังไม่สามารถอธิบายได้ การวิจัยในอนาคตจึงควรพิจารณา การนำตัวแปรอื่นๆ ที่อาจมีผลกระทบเข้าร่วมในการวิเคราะห์ เช่น ดัชนีพลาสติก (Plasticity Index), ปริมาณส่วนละเอียด หรือคุณสมบัติทางเคมีของดิน เพื่อเพิ่มความสามารถในการอธิบายและความแม่นยำของแบบจำลองให้สูงขึ้น ประการที่สอง การศึกษานี้ใช้แบบจำลอง Machine Learning เพียงประเภทเดียว งานวิจัยในอนาคตอาจทำการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองหลายๆ ประเภท เช่น การเปรียบเทียบกับแบบจำลอง ANN ที่ใช้ในงานวิจัยของ Li, J., et al. (2022) [10] หรือ Support Vector Machines เพื่อค้นหาแบบจำลองที่เหมาะสมที่สุดสำหรับชุดข้อมูลประเภทนี้โดยเฉพาะ

สรุปผลการวิจัย

การวิจัยนี้ประสบความสำเร็จในการพัฒนาแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) เพื่อพยากรณ์ค่า DD โดยใช้ตัวแปร MC เป็นตัวแปรพยากรณ์หลักเพียงตัวเดียว แบบจำลองสุดท้ายที่ประกอบด้วยจำนวนต้นไม้ 296 ต้น มีประสิทธิภาพในการทำนายบนชุดข้อมูลทดสอบ (Test set) ในระดับที่น่าพอใจ โดยสามารถอธิบายความผันแปรของข้อมูลได้ 48.73% (R-squared) และมีความคลาดเคลื่อนเฉลี่ย 12.14% (MAPE) จุดเด่นที่สำคัญคือแบบจำลองมีความน่าเชื่อถือสูง สามารถสรุปผลไปยังข้อมูลใหม่ได้ดีโดยไม่เกิดภาวะการเรียนรู้ที่มากเกินไป (Overfitting) ซึ่งบ่งชี้ว่าแบบจำลองได้เรียนรู้รูปแบบความสัมพันธ์ที่แท้จริงของข้อมูล

ผลการศึกษาชี้ให้เห็นอย่างชัดเจนว่า ความสัมพันธ์ระหว่างตัวแปร MC และค่าพยากรณ์ DD มีลักษณะที่ไม่ใช่เชิงเส้นตรง (Non-linear) ซึ่งแบบจำลองที่พัฒนาขึ้นสามารถจับรูปแบบความสัมพันธ์ที่ซับซ้อนนี้ได้อย่างมีประสิทธิภาพ อันเป็นข้อได้เปรียบเหนือกว่าวิธีการทางสถิติแบบดั้งเดิม โดยสรุป การศึกษานี้ไม่เพียงแต่ยืนยันว่า MC เป็นปัจจัยที่มีอิทธิพลสูงสุดต่อการทำนาย แต่ยัง

นำเสนอแบบจำลองที่มีศักยภาพเป็นเครื่องมือช่วยในการพยากรณ์ที่เชื่อถือได้ ซึ่งสามารถพัฒนาต่อยอดเพื่อเพิ่มความแม่นยำในอนาคตได้โดยการพิจารณาตัวแปรอื่นๆ ที่เกี่ยวข้องร่วมในการวิเคราะห์ต่อไป

ข้อเสนอแนะ

เพื่อให้การพัฒนาแบบจำลองการพยากรณ์ในอนาคตมีความแม่นยำและครอบคลุมมากยิ่งขึ้น งานวิจัยในอนาคตควรพิจารณาการนำเข้าตัวแปรพยากรณ์เพิ่มเติม นอกเหนือจากตัวแปร MC เพียงอย่างเดียว เช่น คุณสมบัติพื้นฐานทางปฐพีกลศาสตร์ (Atterberg limits, การกระจายขนาดคละ) เพื่อเพิ่มความสามารถในการอธิบายความผันแปรของข้อมูลและเพิ่มค่า R-squared ให้สูงขึ้น นอกจากนี้ ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของแบบจำลองที่พัฒนานี้กับเทคนิคการเรียนรู้ของเครื่องประเภทอื่นๆ เช่น โครงข่ายประสาทเทียม (Artificial Neural Networks) และควรมีการทดสอบความทนทานของแบบจำลอง (Robustness) กับชุดข้อมูลที่มีความหลากหลายมากขึ้นทั้งในแง่ของชนิดดินและสภาพแวดล้อม เพื่อยืนยันถึงความสามารถในการนำไปประยุกต์ใช้งานได้ในวงกว้างและเพิ่มความน่าเชื่อถือของแบบจำลองต่อไป

เอกสารอ้างอิง

[1] Zhang, X., et al. (2021). Variability in soil composition and its impact on compaction test results. Journal of Geotechnical Engineering.
[2] Mahajan, R., et al. (2020). Impact of environmental conditions on soil compaction testing. International Journal of Civil Engineering.
[3] Miller, T., et al. (2019). Operator-induced errors in compaction testing and standardization. Geotechnical Testing Journal.
[4] Zhou, X., et al. (2021). Mean-based methods for generating standard compaction lines. Journal of Soil Mechanics.
[5] Chen, W., et al. (2020). Polynomial fitting techniques in soil compaction analysis. International Journal of Civil Engineering.
[6] Gupta, N., et al. (2019). Mathematical modeling of compaction curves using polynomial regression. Soil and Foundation Journal.
[7] Singh, P., et al. (2021). Standard line generation for compaction curves: Statistical approaches. Soil Mechanics Journal.
[8] Kim, H., et al. (2020). Computational methods for representing compaction curve variability. Journal of Applied Soil Mechanics.
[9] Arshad, N., et al. (2019). Unifying compaction curves: Challenges and techniques. Engineering Geology.
[10] Li, J., et al. (2022). ANN-based prediction of soil compaction properties. Computers and Geotechnics.
[11] Alam, M., et al. (2021). Machine learning approaches for soil mechanics. Artificial Intelligence in Civil Engineering.
[12] Kumar, R., et al. (2020). Nonlinear modeling of soil compaction using artificial neural networks. International Journal of Geotechnical Engineering.
[13] Agarap, A. F. (2018). Deep Learning Using Rectified Linear Units (ReLU). arXiv preprint arXiv:1803.08375.