

پیش بینی نرخ نفوذ ماشین حفر تونل در سنگ آهک کرتاسه جنوب تهران به روش یادگیری ماشین

آمنه دردشتی^۱، رسول اجل لوئیان^{۲*}، جمال رستمی^۳، جعفر حسن پور^۴، علیرضا سلیمی^۵

پذیرش مقاله: ۱۴۰۲/۰۵/۳۱

دریافت مقاله: ۱۴۰۱/۰۵/۰۱

چکیده

تاکنون تحقیقات زیادی در زمینه پیش بینی عملکرد ماشین حفر تونل در سنگ صورت گرفته است ولی مطالعات قبلی اغلب در شرایطی صورت گرفته اند که خردایش (Chipping) فرآیند غالب در حفاری بوده و مدل های پیش بینی عملکرد ماشین نیز در این شرایط توسعه یافته اند. در مواقعی که به هر دلیلی نیروی تراست دیسک کاتر برای نفوذ در سنگ کافی نباشد و خردایش کامل نشود، کارایی مدل های موجود کم می شود و بایستی مدل های جدیدی برای شرایط آسیاب کاری (Grinding) ارائه گردد. در این مطالعه داده های بدست آمده در حین اجرای تونل پروژه توسعه جنوبی خط ۶ مترو تهران، که بخشی از مسیر آن در واحدهای سنگ آهک کرتاسه قرار داشته و به دلیل عدم کفایت نیروی تراست دیسک کاترها، آسیاب کاری فرایند غالب حفاری بوده است، مورد ارزیابی قرار گرفته است. هدف پیاده سازی الگوریتم های مختلف یادگیری ماشین برای پیش بینی پارامتر نفوذ کله حفاری در هر دور چرخش در شرایط آسیاب کاری در سنگ آهک مورد مطالعه است. به همین منظور الگوریتم های مختلف رگرسیون خطی و غیرخطی پیاده سازی و در نهایت نتایج با هم مقایسه شده است. از آنجاکه تعدادی از این الگوریتم ها جزء جدیدترین رویکردهای یادگیری ماشین هستند، خطای پیش بینی بسیار کم تر از روش های مرسوم رگرسیون می باشد.

کلید واژه ها: آنالیز رگرسیون، آسیاب کاری، عملکرد ماشین حفر تونل، نرخ نفوذ، یادگیری ماشین.

۱. دانشجوی دکتری زمین شناسی مهندسی، دانشگاه اصفهان، اصفهان

۲. عضو هیات علمی گروه زمین شناسی، دانشگاه اصفهان، اصفهان rasajl@sci.ui.ac.ir

۳. عضو هیات علمی دانشکده معدن، Colorado School of Mines، آمریکا

۴. عضو هیات علمی گروه زمین شناسی، دانشگاه تهران، تهران

۵. دکترای مهندسی ژئوتکنیک/تونل زنی و مهندسی سنگ در ZETCON Ingenieur GmbH، آلمان

* مسئول مکاتبات

۱. مقدمه

ماشین تمام‌مقطع حفر تونل (TBM) یکی از متداول‌ترین تجهیزات مورد استفاده در صنعت تونل‌سازی است و در شرایط زمین‌شناسی مختلف مورد استفاده قرار می‌گیرد. از ماشین‌های حفر تونل در شرایط مختلف، برای تونل‌های با سایز و کاربردهای مختلف استفاده می‌شود. تونل‌سازی با استفاده از این ماشین‌ها سرعت بالاتر، نیاز به پرسنل کمتر، محیط کار امن‌تر و هزینه کمتری را به همراه دارد. از سوی دیگر، پیش‌بینی عملکرد این ماشین‌ها در شرایط مختلف، یک بخش مهم در طراحی تونل و انتخاب نوع ماشین و مشخصات مرتبط با آن است. عملکرد ماشین حفر تونل از طریق سه پارامتر نرخ پیشروی روزانه (Advance rate)، نرخ نفوذ (Penetration rate) و ضریب بهره‌وری (Utilization factor) ماشین محاسبه می‌شود. میزان حفاری TBM در سنگ با نرخ نفوذ (ROP) بیان می‌شود که عبارتست از نسبت طول حفاری به زمان واقعی حفاری در طول یک سیکل حفاری پیوسته که از طریق روابط (۱) و (۲) بدست می‌آید.

$$ROP(m/hr) = \frac{L(m)}{t(hr)} \quad (1)$$

$$PR(mm/rev) = \frac{ROP(m/hr) \cdot 1000}{RPM \cdot 60} \quad (2)$$

در روابط فوق ROP نرخ نفوذ برحسب (m/hr) ، L طول حفاری بر حسب متر، t زمان خالص حفاری بر حسب ساعت، PR نفوذ کله حفاری در هر دور چرخش برحسب (mm/rev) و RPM سرعت چرخش کله حفاری بر حسب (rev/min) می‌باشد. از زمانی که ماشین حفر تونل در دهه ۱۹۵۰ معرفی شد، مهندسين طراح و پیمانکاران با مشکل دقت پیش‌بینی عملکرد ماشین برای یک پروژه روبرو بودند. اگرچه در تونل‌سازی مکانیزه پیشرفت‌های زیادی صورت گرفته است، اما تنوع و گستردگی استفاده از این ماشین‌ها در شرایط زمین‌شناسی مختلف، باعث شده است پیش‌بینی عملکرد آن‌ها دشوار باشد و عدم قطعیت‌های زیادی در مدل‌های توسعه‌یافته وجود داشته باشد. مهمترین پارامترهای تأثیرگذار بر عملکرد ماشین به دو گروه عمده پارامترهای اپراتوری ماشین و

پارامترهای زمین‌شناسی تقسیم می‌شوند (حسن‌پور و همکاران، ۱۴۰۲). درک صحیح این پارامترها و نحوه محاسبه آن‌ها در شناخت دقیق پارامترهای مؤثر بر عملکرد ماشین بسیار ضروری است. روش‌های محاسبه نرخ پیشروی ماشین که در صنعت به کار می‌روند، دارای نقاط قوت و ضعفی هستند که امروزه تلاش می‌شود دقت آن‌ها تا حد ممکن بالا رود و کمبودهای آنها پوشش داده شود. به منظور کاربرد موفقیت‌آمیز ماشین حفاری در یک پروژه تونل‌سازی نیاز به توسعه تحقیقات در هر دو زمینه اثر شرایط زمین‌شناسی و طراحی جزئیات ماشین وجود دارد.

تاکنون تحقیقات فراوانی جهت پیش‌بینی نرخ نفوذ و عملکرد ماشین و بررسی ارتباط آن با پارامترهای زمین‌شناسی و راهبری ماشین صورت پذیرفته است. اگرچه برخی از مدل‌های ارائه شده تا حدود زیادی موفق بوده‌اند، اما اغلب دارای محدودیت‌هایی در استفاده هستند. این مدل‌ها شامل دو گروه: (۱) مدل‌های تئوری که بر اساس بررسی مراحل خرد شدن سنگ در محیط آزمایشگاه توسعه یافته‌اند و (۲) مدل‌های تجربی که با استفاده از داده‌های واقعی از عملکرد ماشین به دست آمده‌اند، می‌باشند. مزیت روش‌های تجربی این است که به طور طبیعی تمام تأثیرات ناشی از زمین و ماشین را در نظر می‌گیرند، اما با این حال، رشد فناوری و ارائه‌ی ماشین‌هایی با قابلیت‌های متفاوت، استفاده از این روش‌ها را محدود می‌سازد. ضمن این‌که چون مدل‌های ارائه شده برای شرایط خاص زمین پیشنهاد شده‌اند، قابل تعمیم به تمام شرایط زمین نیستند و همبستگی به‌دست‌آمده به‌طورکامل به خصوصیات TBM و شرایط زمین مورد مطالعه وابسته است. با این حال روش‌های تجربی نسبت به روش‌های گران و زمان‌بر آزمایشگاهی مناسب‌تر بوده و تطابق بیشتری با واقعیت دارند.

در سال‌های اخیر مدل‌سازی‌های کامپیوتری و روش‌های یادگیری ماشین (Machine Learning) نیز توسط برخی محققان جهت توسعه مدل‌های تجربی پیش‌بینی نرخ نفوذ و عملکرد ماشین به‌کارگرفته شده‌اند (Grima et al., 2000; Koopialipour et al., 2019; Yagiz and Karahan, 2011).

بایستی مدل‌های جدیدی ارائه گردد. خردایش (Chipping) فرایند اصلی حفاری سنگ در ماشین‌های تونل‌بری در سنگ سخت است و مطالعات نشان می‌دهند که نفوذ اولیه باید از $4-5 \text{ mm/rev}$ تجاوز کند تا خردایش موثر اتفاق افتد (Pourhashemi et al., 2021). زمانی که ترک‌های شعاعی حاصل از حفاری با دیسک‌کاترها در سنگ به میزان کافی گسترش پیدا نکنند، فرآیند آسیاب‌کاری (Grinding) صورت گرفته و این امر معمولاً به‌خاطر کافی نبودن نیروی دیسک‌کاترها و یا سختی بالای سنگ و نفوذ اولیه ناکافی به‌وقوع می‌پیوندد. در این شرایط مقدار زیادی پودرسنگ تولید می‌شود و یا بلوک‌های سنگی تشکیل می‌شوند و در هر دو صورت راندمان حفاری بسیار پایین می‌آید و معمولاً به دلیل ایجاد برآمدگی‌ها (ridges) و تماس با کله‌حفاری، به کله‌حفار و دیسک‌کاترهای نصب شده بر روی آن آسیب زیادی وارد می‌شود. نیروی تراست مورد نیاز (Fn) برای عبور از حالت آسیاب‌کاری (Grinding) در سنگ و شروع خردایش (Chipping)، نیروی تراست بحرانی (Critical Thrust) نامیده می‌شود. مطالعات پیشین نشان می‌دهند که مقاومت‌سنگ، یک پارامتر مهم برای تعیین نیروی تراست بحرانی در سنگ‌های مختلف است (Robbins, 1970). تاکنون مطالعاتی در زمینه پدیده آسیاب‌کاری (Grinding) و خردایش (Chipping) و مرزهای بین آن‌ها در سنگ‌های مختلف انجام شده است. برخی از مهم‌ترین تحقیقات در این زمینه، نتایج مطالعات رابینز (Robbins, 1970)، برولند (Bruland, 1998)، گیرمیشید (Girmscheid, 2008)، گهرینگ (Gehring, 2009)، ویلنیو (Villeneuve, 2008, 2012, 2017) و تورو (Thuro, 2017) می‌باشند. در تمامی مطالعات صورت گرفته رابطه بین نفوذ کله‌حفار با نیروی تراست موردنیاز بررسی و روابطی ارائه شده است.

یکی از پروژه‌هایی که در آن خردایش سنگ زیر دیسک‌کاتر کامل‌نشده و فرآیند آسیاب‌کاری غالب بوده است، تونل توسعه جنوبی خط ۶ متروی تهران (SEL6) بوده که به دلیل الزامات طراحی ماشین و کافی نبودن نیروی تراست،

Salimi et al., 2016, 2019; Jahed Armaghani et al., (2017; Zhou et al., 2020).

در این سال‌ها استفاده از تکنیک‌های هوش مصنوعی (Artificial Intelligence (AI)) در تخمین عملکرد TBM بسیار مورد توجه قرار گرفته است. در مواردی که رابطه‌ی مشخصی میان پارامترهای ورودی و خروجی وجود نداشته و ارتباط میان آن‌ها توسط روش‌های مرسوم قابل شناسایی نیست، استفاده از روش‌های هوش مصنوعی به سبب عدم نیاز به فرمول‌بندی، پویا بودن و همچنین امکان لحاظ کردن تأثیر پارامترهای مؤثرتر، ایده‌ی خوبی برای تخمین و مدل‌سازی به‌شمار می‌رود. طی تحقیقات اخیر و با مقایسه‌ی مقادیر پیش‌بینی‌شده با استفاده از هوش مصنوعی و رگرسیون چندمتغیره‌ی غیرخطی با مقادیر واقعی نرخ نفوذ، این نتیجه حاصل شده است که نتایج هوش مصنوعی نسبت به رگرسیون غیرخطی به واقعیت نزدیک‌تر است. البته باید توجه داشت، برخی تکنیک‌های هوش مصنوعی مانند ANN یا SVR که به عنوان تکنیک‌های قدرتمندی در تخمین عملکرد TBM هستند، به راحتی قابل تفسیر و توضیح نیستند و به عنوان روش‌های جعبه سیاه یا Black Box شناخته می‌شوند چراکه قوانین یا الگوهای ضمنی ساخت مدل بسیار پیچیده‌اند. در واقع هوش مصنوعی با توجه به قابلیت‌هایش، می‌تواند راه‌های ساده‌تر، کم هزینه‌تر و نتایج دقیق‌تری را پیش روی مهندسين قرار دهد. طبیعت انعطاف پذیر تکنیک‌های AI آن‌ها را به ابزارهای قدرتمندی در ارزیابی و حل مسائل مهندسی بخصوص مسائل پیچیده و غیرخطی تبدیل کرده است. این مطالعات، همبستگی زیادی بین نرخ‌های پیش‌بینی شده و مقادیر واقعی عملکرد ماشین نشان می‌دهند (He et al., 2022). نکته مهم آن است که مطالعات قبلی در زمینه پیش‌بینی عملکرد ماشین، اغلب در شرایطی صورت گرفته‌اند که خردایش سنگ (Chipping) فرآیند غالب در حفاری با دیسک‌کاتر بوده و مدل‌های پیش‌بینی عملکرد TBM در سنگ نیز در این شرایط توسعه یافته‌اند. در مواقعی که به هر دلیلی نیروی تراست دیسک‌کاتر برای نفوذ در سنگ کافی نباشد و خردایش کامل نشود، کارایی مدل‌های موجود کم می‌شود و

متوسط لایه تا توده‌های با رنگ قهوه‌ای روشن و خاکستری می‌باشند. همانطور که اشاره شد و در برش زمین‌شناسی مسیر تونل نیز دیده می‌شود (شکل ۱)، مسیر تونل از دو واحد زمین‌شناسی متفاوت شامل آبرفت‌های کواترنری تهران (متشکل از خاک‌های ریزدانه) و سنگ‌آهک کرتاسه عبور می‌کند. همانطور که در شکل ۱ مشخص است، در محل تماس لایه‌های سنگ‌آهک و آبرفت نیز زون انتقالی شامل خاک آمیخته با قطعات هوازده سنگ‌های آهکی وجود دارد (Firouzei et al., 2019). در ابتدای تونل SEL6 حفاری در واحد خاک ریزدانه آبرفت تهران صورت گرفته و بعد از حدود ۸۰۰ متر از ورودی (Portal) دولت‌آباد، تونل به سنگ‌آهک متوسط تا ضخیم لایه کرتاسه برخورد کرده که حدود ۱۲۰۰ متر ادامه دارد و پس از آن حفاری در آبرفت ریزدانه تهران ادامه می‌یابد (مهندسین مشاور رهساز طرح، ۱۳۹۵). شکل ۲ رخنمون توده سنگ آهکی مسیر تونل و سنگ‌آهک مورد مطالعه داخل تونل حین حفاری را نشان می‌دهد که طی مطالعات برداشت شده است.

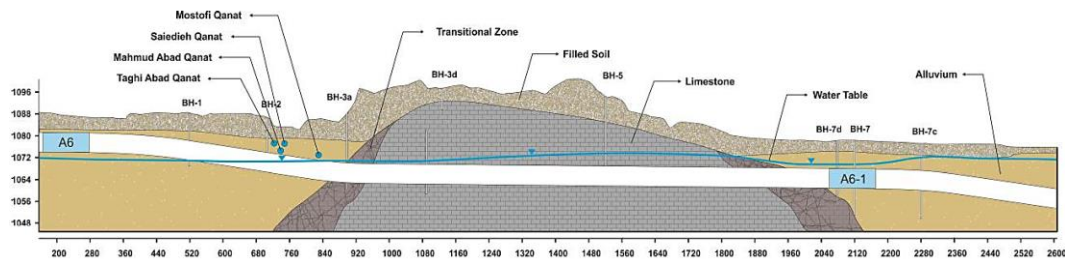
این تونل با استفاده از یک ماشین TBM "تبادل فشار زمین" یا EPB ساخت شرکت هرکنشت آلمان، با قطر ۹/۱۹ متر حفاری و سگمنت‌گذاری شده است. این ماشین در اصل برای حفاری در شرایط زمین نرم با درصد ریزدانه بالا طراحی شده است. مشخصات فنی این ماشین در جدول ۱ ارائه شده است. برای حفاری بخش سنگی تونل، از ۲۶ دیسک‌کاتر دوتایی به قطر ۱۷ اینچ که بر روی کله‌حفار به جای ابزار برش زمین نرم (درگ‌بیت‌ها) نصب شده‌اند، استفاده شده است. این دیسک‌کاترها در مجموع ۵۲ لبه برشی داشته و ۵۲ اثر بر روی سینه‌کار تونل باقی می‌گذارند (مهندسین مشاور رهساز طرح، ۱۳۹۵).

مشخصات واحدهای سنگی و پارامترهای ژئومکانیکی توده‌سنگ‌های مسیر تونل که در رخنمون‌های سطحی و سینه‌کار تونل مطالعه شده‌اند، در جدول ۲ ارائه شده است.

آسیاب‌کاری رخ داده است. در این پروژه، در ابتدا بر اساس شرایط زمین‌شناسی مسیر اولیه تونل که پیش‌بینی می‌شد به طور کامل در زمین نرم است، نوع ماشین تعیین و ابزارهای برش انتخاب شد. در ادامه مطالعات زمین‌شناسی مهندسی، با توجه به ثابت بودن مسیر تونل به دلیل الزامات ترافیک شهری و بهره‌برداری، مشخص شد که حدود ۱۲۰۰ متر از طول مسیر تونل در سنگ‌آهک کرتاسه خواهد بود. این در حالی است که ماشین انتخابی اصولاً برای حفاری در سنگ سخت طراحی نشده و پیش‌بینی می‌شد که حفاری با این ماشین با چالش‌هایی از جمله عدم کفایت بار اعمالی کاترها در سنگ روبه‌رو باشد. به همین دلیل داده‌های حفاری این تونل جهت بررسی رابطه بین میزان نفوذ کله‌حفار و خصوصیات زمین‌شناسی مهندسی سنگ‌آهک مورد مطالعه در شرایط آسیاب‌کاری مناسب می‌باشد. در این مقاله با استفاده از روش‌های آماری و یادگیری ماشین، یک مدل جدید جهت پیش‌بینی عملکرد TBM بر اساس داده‌های واقعی عملکرد ماشین و خصوصیات زمین‌شناسی مهندسی توده‌سنگ، برای شرایط آسیاب‌کاری در سنگ‌آهک جنوب تهران ارائه می‌گردد.

۲. معرفی پروژه

تونل SEL6 با طول ۶ کیلومتر و قطر ۹/۲ متر بخشی از توسعه جنوبی خط ۶ متروی تهران است. در بخشی از مسیر این تونل سنگ‌های آهکی به سن کرتاسه در میان آبرفت تهران بیرون‌زدگی دارند. بر اساس مطالعات طریق‌زلی و همکاران (Tarigh Azali et al., 2013) دشت تهران عمدتاً از سازندهای کواترنر تشکیل شده که اغلب نتیجه فرسایش و رسوب مجدد رسوبات قبلی است و شامل رسوبات آبرفتی و نهشته‌های رودخانه‌ای طبقه‌بندی نشده است که به سمت جنوب گسترش یافته است. کوه بی‌بی‌شهربانو به صورت ستیغ‌هایی در جنوب شرق تهران و در حاشیه شرقی شهری نمایان است که تپه‌ها در ادامه کوه بی‌بی‌شهربانو به سمت داخل شهر کشیده شده است. تپه‌ها از جنس دولومیت و سنگ‌آهک به سن تریاس و کرتاسه بصورت توده‌سنگ‌های



شکل ۱. مقطع زمین شناسی تونل SEL6 خط ۶ مترو تهران (Firouzei et al., 2019)

جدول ۱. مشخصات فنی ماشین مورد استفاده در تونل توسعه جنوبی خط ۶ متروی تهران

Descriptions		Specifications
Cutterhead	TBM diameter	9.19 m
	Number of cutting tools	52
	Disc cutter diameter	17 (in.)
	Number of central and peripheral D. cutting tools	26
Driver system	Electromotor power	2400 (kW)
	Ave. Cutterhead rotation speed	0.3-8 RPM
	Max. cutterhead torque	17197 (kN.m)
Thrust system	Number of thrust cylinders	33
	Max. total Thrust	30000 (kN)
Screw conveyor	Ave. screw conveyor torque	0-22 (kN.m)
	Ave. screw conveyor rotation speed	2.3 RPM



(ب)

(الف)

شکل ۲. الف) رخنمون توده سنگ آهکی مسیر تونل ب) سنگ آهک داخل تونل SEL6

جدول ۲. پارامترهای ژئومکانیکی توده سنگ‌های مسیر تونل SEL6

Engineering Geological units	Geotechnical characteristics				Rock mass geomechanical parameters				
	lithology	Structure	Strength degree	Stability condition	BRMR	UCS (MPa)	GSI	RQD	Q
RT1	Limestone	Layered, Jointed	Strong, very Strong	Local structural instability	60-75	80-120	60-65	90-100	6.27
RT2	Limestone	Weathered	Strong	Unstable	45-55	50-80	40-45	60-80	1.85

BRMR: سیستم طبقه‌بندی سنگ RMR پایه؛ UCS: مقاومت تراکمی تک‌محوره؛ GSI: شاخص مقاومت زمین‌شناسی؛ RQD: شاخص کیفیت سنگ؛ Q: سیستم طبقه‌بندی سنگ

۳. مواد و روش‌ها

در هر پروژه تونل‌سازی، خصوصیات ماشین حفاری براساس شرایط زمین‌شناسی غالب مسیر تونل طراحی می‌شود. همچنین زمین‌شناسی مسیر تا حدود زیادی تعیین‌کننده نوع و موقعیت نصب ابزارهای برش روی کله‌حفار است. در این پروژه، در ابتدا بر اساس شرایط زمین‌شناسی مسیر اولیه تونل (که پیش‌بینی می‌شد به طور کامل در زمین نرم است)، روش اجرا، نوع ماشین و ابزارهای برش انتخاب شد. در ادامه مطالعات زمین‌شناسی مهندسی، با توجه به ثابت بودن مسیر تونل به دلیل الزامات اجرا و بهره برداری، مشخص شد که حدود ۱۲۰۰ متر از طول مسیر در سنگ آهک خواهد بود. این در حالی بود که ماشین انتخابی برای حفاری در سنگ سخت طراحی نشده بود. بنابراین به دلایلی که در ادامه تشریح می‌گردد، پیش‌بینی می‌شد که حفاری با این ماشین در سنگ آهک با چالش‌هایی بزرگ از جمله کافی نبودن بار اعمالی کاترها روبرو باشد.

۳-۱. چالش حفاری در سنگ در پروژه

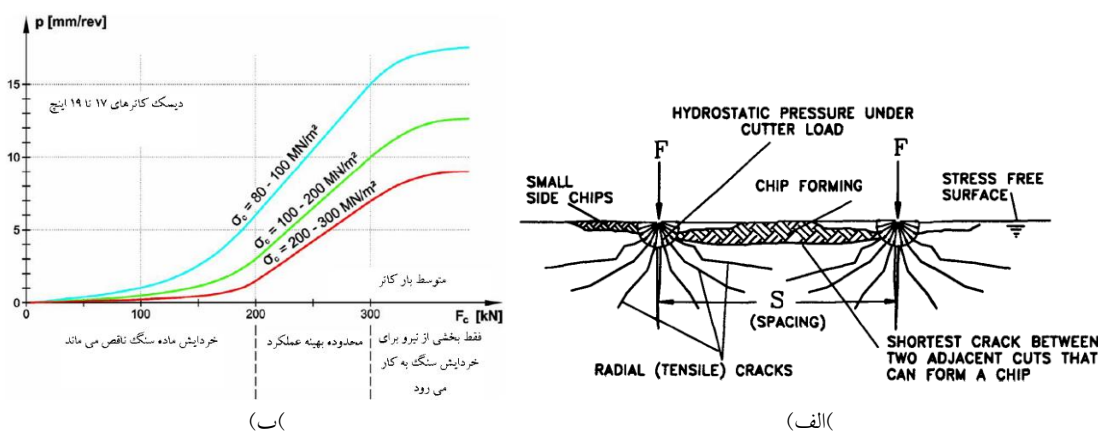
به طور کلی در فرآیند حفاری با ماشین TBM، خردایش سنگ در اثر نفوذ دیسک‌کاتر به دو مرحله زیر قابل تقسیم می‌باشد (Rostami and Ozdemir, 1993):

- تشکیل میدان تنش و تشکیل زون خردشده
- تشکیل تراشه‌های سطحی و تشکیل شیار

در مرحله اول دیسک‌کاتر در حال چرخش، تحت تأثیر نیروی پیشران در سنگ شروع به نفوذ کرده و به دلیل تنش‌های بسیار بالا، زونی از مواد خردشده در زیر دیسک‌کاتر تشکیل می‌گردد. ابعاد این زون با افزایش فشار گسترش می‌یابد تا هنگامی که مقدار فشار به حد بحرانی می‌رسد. در این هنگام کرنش‌های کششی ایجاد شده در زون خردشده، سبب تشکیل و گسترش ترک‌های کششی کوچک و بزرگ در محل تماس دیسک‌کاتر با سطح سنگ می‌شوند. گسترش ترک‌ها تا جایی ادامه می‌یابد که تنش‌های کششی موجود از مقاومت کششی

سنگ کمتر گردد. این مرحله را می‌توان مرحله نفوذ یا فرو رفتن دیسک‌کاتر نام‌گذاری نمود. در مرحله دوم اگر تنش‌های توسعه یافته در زون خردشده به اندازه کافی بزرگ باشند، ترک‌های تشکیل شده میان دو دیسک‌کاتر مجاور به سوی یکدیگر گسترش یافته و به هم می‌پیوندند. در نتیجه همان‌طور که در شکل ۳ الف نشان داده شده است، تراشه‌های بزرگی بین دو دیسک‌کاتر مجاور تشکیل می‌گردد. با تشکیل تراشه‌ها و جدا شدن آن‌ها از سینه‌کار تونل، فرایند خردایش سنگ تکمیل گشته و کله‌حفار در سنگ پیشروی می‌نماید.

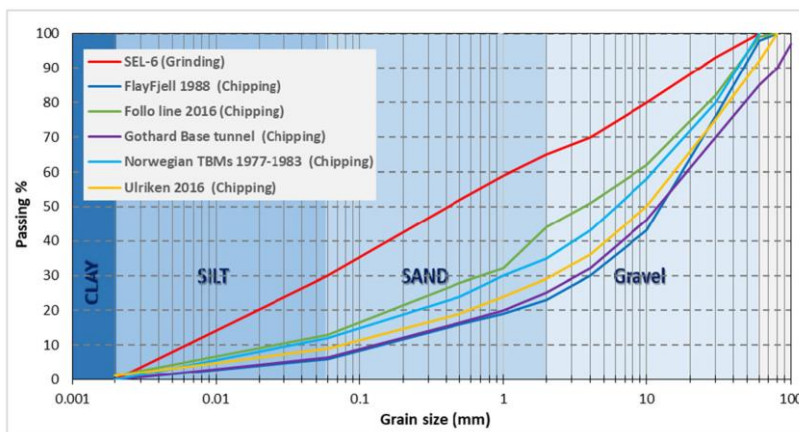
رابطه بین مقاومت سنگ‌بکر، نیروی تراست دیسک‌کاتر و مقدار نفوذ (mm/rev) توسط بسیاری از محققان از جمله فرنزل و همکاران (Frenzel et al., 2008) مطالعه شده‌است. مطالعات این محققان نشان می‌دهد که باید نفوذ اولیه کله‌حفار بیش از ۵-۴ میلیمتر در هر چرخش باشد تا فرآیند خردایش و تشکیل تراشه کامل باشد (Pourhashemi et al., 2021). شکل ۳ ب وابستگی فرایند خردایش سنگ و مشخصاً پارامتر نفوذ کله‌حفار (mm/rev) را به بار دیسک‌کاتر برای سنگ‌هایی با مقاومت تک‌محوری متفاوت و برای دیسک‌کاترهایی با قطر ۱۷ تا ۱۹ اینچ را نشان می‌دهد (Girmschid, 2005). بر اساس این شکل اگر بار دیسک‌کاتر اندک باشد، تنش فشاری ایجاد شده توسط دیسک‌کاتر برای ایجاد تنش‌های کششی که سبب تشکیل ترک‌ها و در نهایت تشکیل تراشه‌ها می‌شوند، کافی نخواهد بود، بنابراین در این حالت فرایند خردایش سنگ ناقص خواهد ماند. مطابق با شکل ۳ ب محدوده بهینه بار دیسک‌کاترها برای سنگ‌هایی با مقاومت ۸۰ تا ۳۰۰ مگاپاسکال، ۲۰۰ تا ۳۰۰ کیلونیوتن برای هر حلقه دیسک‌کاتر است. بنابراین، مقدار نیروی پیشران در تکمیل فرایند تشکیل تراشه‌ها بسیار حائز اهمیت بوده و عدم کفایت این نیرو می‌تواند سبب وقوع پدیده آسیاب‌کاری شود که در کاهش عملکرد ماشین و افزایش سایش ابزار برش بسیار مؤثر خواهد بود.



شکل ۳. الف) مدلی شماتیک از فرایند خردایش و تشکیل تراشه در زیر دیسک کاتر (Rostami and Ozdemir, 1993) ب) مقادیر نفوذ دیسک کاتر بر حسب بار دیسک کاتر و مقاومت فشاری ماده سنگ (Girmschid, 2005)

ظرفیت تحمل بار دیسک کاترهای نصب شده، حداکثر نیروی مجاز قابل اعمال بر روی هر حلقه ۱۳۵ کیلو نیوتن تعیین گردید که مطابق شکل ۳ب برای ایجاد نفوذ اولیه بیش از ۵ میلیمتر در هر دور، در سنگ‌آهک با ویژگی‌های مندرج در جدول ۲ کافی نیست. در نتیجه هنگام حفاری با این ماشین، در اغلب موارد فرآیند خردایش کامل نبوده و آسیاب‌کاری پدیده غالب بوده است. در این شرایط توزیع دانه‌بندی ذرات حاصل از حفاری نسبت به حالت معمول بسیار متفاوت است. شکل ۴ مقایسه توزیع دانه‌بندی ذرات حاصل از حفاری پروژه‌های مختلف TBM در اروپا را با توزیع دانه‌بندی ذرات حاصل از حفاری در پروژه SEL6 که در آن فرآیند آسیاب‌کاری غالب بوده است، نشان می‌دهد. واضح است که مواد حفاری در این پروژه بسیار ریزتر از بقیه پروژه‌هاست.

در این پروژه برای حفاری در سنگ‌آهک با توجه به UCS توده‌سنگ، دیسک کاترهای دوتایی به قطر ۱۷ اینچ به جای ابزار برش زمین نرم (درگ بیت‌ها) نصب شدند. یک دیسک کاتر دوتایی شامل دو حلقه دیسک با یک بلبرینگ (Bearing) می‌باشد. مطابق معیار طراحی برای کاتر ۱۷ اینچ، هر بلبرینگ می‌تواند تا نیروی ۲۷۰ kN را تأمین کند. بنابراین در طول حفاری اپراتور نمی‌تواند نیرویی بیش از این مقدار را روی هر حلقه دیسک اعمال کند. به بیان دیگر هر حلقه نمی‌تواند نیرویی بیش از ۱۳۵ kN بر سنگ وارد کند که برای ایجاد نفوذ بیش از ۵ میلی‌متر کافی نیست، بنابراین این پروژه فرصتی برای پیدا کردن روابط بین نفوذ ماشین و خصوصیات زمین‌شناسی مهندسی در شرایط آسیاب‌کاری به دست می‌دهد. به این ترتیب در ماشین مورد استفاده در پروژه با توجه به



شکل ۴. مقایسه توزیع دانه‌بندی ذرات حاصل از حفاری در پروژه‌های مختلف در شرایط خردایش عادی و در تونل SEL6 در شرایط آسیاب‌کاری (Pourhashemi et al., 2021)

۳-۲. توسعه پایگاه داده

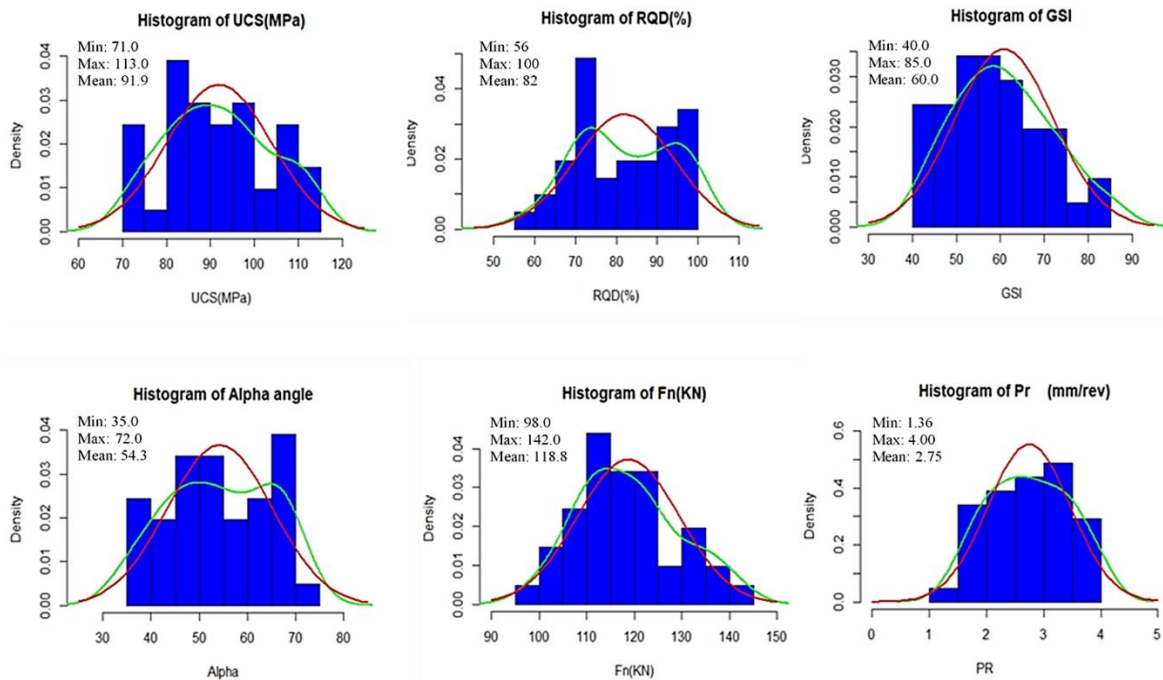
اولین قدم برای ساخت مدل جهت پیش‌بینی عملکرد TBM، انتخاب پارامترهای ورودی مناسب است. این ورودی‌ها به دو دسته خصوصیات توده‌سنگ و خصوصیات ماشین تقسیم می‌شوند (حسن‌پور و همکاران، ۱۴۰۲). مرحله انتخاب پارامترهای ورودی مهم‌ترین و بحرانی‌ترین قسمت در مدل‌سازی است. در این مرحله بر اساس داده‌های جمع‌آوری شده از پروژه، داده‌ها آماده‌سازی می‌شوند. این آماده‌سازی شامل آنالیزهای آماری توصیفی، حذف داده‌های نامتعارف (Outlier) به کمک برخی تکنیک‌های یادگیری ماشین، شناسایی داده‌های گم‌شده و غیره می‌باشند که به این ترتیب داده‌ها پردازش شده و از تاثیر منفی آن‌ها بر روی مدل جلوگیری می‌گردد.

داده‌های جمع‌آوری شده شامل ۴۱ دسته داده می‌باشد که برای پیش‌بینی عملکرد TBM در شرایط آسیاب‌کاری به کار می‌رود. پارامترهای عملکرد ماشین شامل نیروی تراست دیسک‌کاتر (Fn)، RPM، گشتاور، میزان نفوذ کله‌حفار (PR)، نرخ نفوذ (ROP) و زمان حفاری برای هر رینگ ۱/۵ متری به صورت میانگین ثبت شده است و از میان این داده‌ها، پارامترهای نفوذ کله‌حفار در هر دور چرخش $(PR)mm/rev$ و نیروی تراست (Fn) در مدل‌سازی مورد استفاده قرار گرفت. پارامترهای زمین‌شناسی مهندسی توده‌سنگ و سنگ‌بکر از منحنی توزیع

دانه‌بندی ذرات حاصل از حفاری، نمونه‌گیری، مغزه‌گیری و آزمون‌های آزمایشگاهی به دست آمده است. پارامترهای لیتولوژی، درجه هوازدگی، نوع و شرایط ناپیوستگی شامل جهت‌یابی، فاصله‌داری، شرایط سطح درزه مطابق استاندارد 1978 (ISRM)، زاویه بین محور تونل و صفحات ناپیوستگی یا لایه‌بندی (α)، RQD و شاخص مقاومت زمین‌شناسی GSI ارزیابی و برآورد شد. برای انجام آزمون‌های آزمایشگاهی شامل بار نقطه‌ای، UCS و آنالیز پتروگرافی، مغزه‌هایی از سینه کار تونل گرفته شده و در نهایت این داده‌ها به ۱/۵ متر رینگ حفاری مربوطه تخصیص یافت. در نهایت از میان پارامترهای زمین‌شناسی مهندسی سنگ UCS، RQD، GSI و زاویه بین محور تونل و صفحات ناپیوستگی (α) در مدل‌سازی مورد استفاده قرار گرفت. انتخاب این پارامترها جهت مدل‌سازی، براساس مطالعه انجام شده بر روی همین پروژه توسط پورهاشمی و همکاران (Pourhashemi et al., 2021) صورت گرفته و سپس با انجام آنالیز همبستگی بین متغیرها، یک دیدگاه کلی پیرامون رابطه بین متغیرها به دست آمد. آنالیز آماری پارامترهای عملکرد ماشین و زمین‌شناسی مهندسی مورد استفاده در مدل‌سازی در جدول ۳ خلاصه شده است. جهت انجام تحلیل آماری توصیفی، توزیع هر یک از پارامترها به کمک نمودار هیستوگرام انجام و در شکل ۵ نشان داده شده است.

جدول ۳. آنالیز آماری پارامترهای مورد استفاده در مدل‌سازی (متغیرهای مستقل و متغیر وابسته PR)

	Alpha angle	UCS (MPa)	RQD	GSI	Fn (kN)	PR (mm/rev)
Min.	35.00	71.00	56.00	40.00	98.0	1.360
1st Qu.	46.00	82.00	73.00	55.00	111.0	2.170
Median	55.00	91.00	82.00	60.00	118.0	2.740
Mean	54.29	91.93	82.00	60.00	118.8	2.748
3rd Qu.	64.00	100.00	94.00	70.00	125.0	3.390
Max.	72.00	113.00	100.00	85.00	142.0	4.000
skewness	-0.12	0.14	-0.06	0.30	0.33	0.02
kurtosis	1.86	2.07	1.91	2.45	2.46	2.03



شکل ۵. نمودار هیستوگرام پارامترهای مورد استفاده در مدل‌سازی (متغیرهای مستقل و متغیر وابسته PR)

۴. مدل‌سازی

وابسته را کشف و توجیه کند. یادگیری نظارت شده یا Supervised Learning که برای پیش‌بینی به کار می‌رود، در این حیطه قرار می‌گیرد.

در این تحقیق با استفاده از متغیرهای مستقل UCS, Fn, GSI, RQD و زاویه α سعی در پیش‌بینی متغیر وابسته یا همان میزان نفوذ کله‌حفار (PR) داریم. از آنجا که متغیر وابسته از نوع داده‌های پیوسته است، الگوریتم‌های رگرسیون برای مدل‌سازی روی این داده‌ها مناسب می‌باشند. برای این کار داده‌ها را به دو بخش Train و Test/Validation تقسیم کرده و روی داده‌های Train مدل‌های مختلف از نسل‌های مختلف الگوریتم‌های ML پیاده‌سازی شده تا به بهترین مدل نزدیک شویم، تا جایی که مسئله تعمیم‌پذیری مدل تامین گردد و مدل روی داده‌هایی که خارج از بازه داده‌های ورودی مدل است، هم خوب عمل کند و به‌درستی پیش‌بینی صورت گیرد. در صورتی که دو فرض رگرسیون خطی برقرار باشد، تعمیم‌پذیری مدل صورت می‌گیرد. این دو فرض عبارتند از: (۱) برای هر ترکیبی از متغیرهای مستقل، مقدار خطا دارای توزیع نرمال با میانگین صفر و انحراف معیار ثابت باشد. (۲)

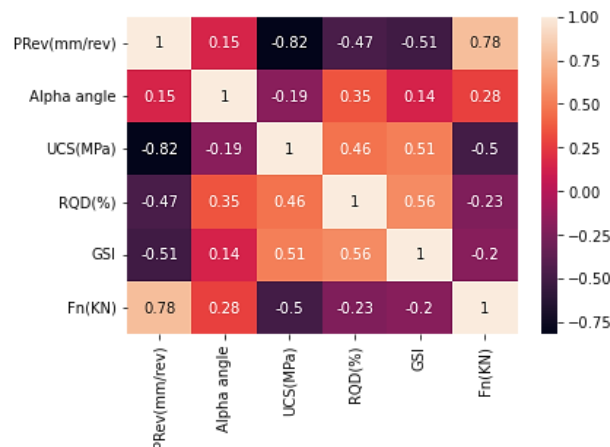
در این مطالعه جهت پیش‌بینی نرخ نفوذ و عملکرد ماشین، روش‌های یادگیری ماشین (Machine Learning) به‌کارگرفته شده است. مدل‌های مبتنی بر AI که تاکنون برای پیش‌بینی عملکرد TBM به کار رفته‌اند، براساس تکنیک‌های یادگیری نظارت‌شده بوده‌اند. در این رویکردها داده‌های ورودی و خروجی به ماشین آموزش داده می‌شوند و سپس ساختار مدل‌های پیش‌بینی‌کننده، ساخته می‌شوند و سپس بهترین مدل برای آنالیز رابطه بین متغیرهای مستقل و متغیر وابسته انتخاب می‌گردد.

در اصول یادگیری ماشین، الگوریتم‌های ریاضی یا Learners تلاش می‌کنند به کمک ترکیبی از متغیرهای مستقل، تابعی را بهینه کنند تا مقدار پیش‌بینی‌شده توسط تابع به مقدار واقعی متغیر وابسته نزدیک شود. نوع متغیر وابسته (پیوسته یا گسسته) تعیین‌کننده نوع الگوریتم‌هایی است که با آن مسائل حل می‌شود. یادگیری الگوریتم یا Learner از طریق داده‌هایی که مشاهده نموده‌است، رخ می‌دهد پس باید داده‌ها در اختیار Learner قرار گرفته تا ارتباط بین متغیرهای مستقل و متغیر

داشت کمترین تاثیر را در روابط رگرسیون داشته باشد. مطالعات قبلی صورت گرفته از جمله مطالعات رابینز (Robbins, 1992)، لوتن (Laughton, 1998) و حسن پور (Hassanpour et al., 2008, 2011, 2015) نیز نشان می دهند، مقاومت تراکمی تک محوره سنگ عامل بسیار مهمی در حفاری در سنگ به وسیله ماشین TBM است. همچنین بایستی به همبستگی بین خود متغیرهای مستقل نیز توجه کرد چرا که ممکن است باعث هم خطی (Multicollinearity) بین متغیرها شود و ضرایب رگرسیون را تحت تاثیر قرار دهد و به این ترتیب ممکن است نتوان اثر خالص متغیرهای مستقل را روی متغیر وابسته تعیین کرد، به طور مثال RQD با GSI همبستگی ۰/۵۶ دارد و این همبستگی ناشی از تاثیر مقدار RQD در تعیین مقدار پارامتر SR (Structure Rating) در شاخص GSI است.

مقادیر خطا از هم مستقل بوده و یا به عبارت دیگر توزیع واریانس خطاها متقارن باشد (Hastie et al., 2009).

از آنجاکه مسئله رگرسیون مطرح است، با آنالیز همبستگی می توان به یک دیدگاه کلی پیرامون رابطه بین متغیرها دست یافت. مقادیر همبستگی بین متغیرها در شکل ۶ ارائه شده است. با ارزیابی همبستگی بین متغیرهای مستقل UCS، Fn، RQD، GSI، α و متغیر وابسته PR، میزان تاثیر هر یک از پارامترها در مقدار PR به دست آمد. همان طور که انتظار می رفت بالاترین همبستگی مربوط به UCS برابر ۰/۸۲ بود و بیشترین تاثیر را روی نفوذ دیسک کاتر داشت و پس از آن Fn با ۰/۷۸+ و سپس GSI با ۰/۵۱- بیشترین همبستگی را با میزان نفوذ نشان دادند. بنابراین باید انتظار داشت که این ۳ پارامتر، بیشترین تاثیر را در رگرسیون داشته باشند. پارامتر α همبستگی کمی با متغیر وابسته PR نشان می دهد و باید انتظار



شکل ۶. مقادیر شدت همبستگی بین متغیرهای مستقل و متغیر وابسته PR

کرد که در حقیقت ناشی از خطاهای اندازه گیری و دیگر خطاهای پارامتریک است. مدل سازی را با مدل کلاسیک رگرسیون خطی در نرم افزار R براساس تمام متغیرهای مستقل موجود شروع کرده و سپس براساس نتایج آزمون t متغیرها بررسی شدند. مقادیر p-value آزمون F کوچکتر از ۰/۰۵ بوده، بنابراین فرض h_0 را رد می کند و حداقل یک متغیر مستقل با متغیر وابسته رابطه خطی دارد. براساس نتایج آزمون t فقط مقادیر p-value متغیرهای UCS و Fn کمتر از ۰/۰۵ و متمایز (significant) شده، یعنی

۴-۱. رگرسیون خطی

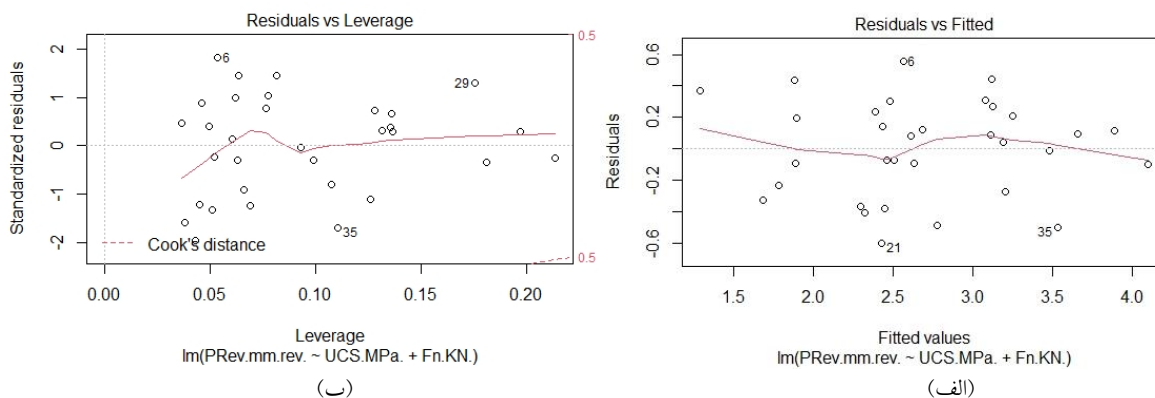
جهت انتخاب متغیرهای مستقلی که در رگرسیون موثرند، ابتدا از رویکردهای کلاسیک رگرسیون مانند رگرسیون بر اساس نتایج آزمون t استفاده می شود ولی گاهی پیش می آید که فرضیات رگرسیون خطی برقرار نیست و باید از روش های جدیدتر مثل Stepwise Reg. یا از رویکردهای قاعده سازی (Regularization) استفاده کرد. این روش ها به دلیل برخی نقاط ضعف آزمون t توسعه یافته اند. در کاربردهای عملی به ندرت می توان رابطه ریاضی دقیق بین دو متغیر را پیش بینی

حذف یک مشاهده نباید تأثیر قابل توجهی روی مدل داشته باشد. نمودار خطاها یا Residuals ($y - \hat{y}$) در برابر Leverage در شکل ۷ نشان می‌دهد که مدل از نظر وجود داده‌های تأثیرگذار، مشکلی ندارد و Cook's distance همه مشاهدات کمتر از یک است.

نکته حائز اهمیت دیگر این‌که زمانی که از رگرسیون برای پیش‌بینی متغیر وابسته بر اساس متغیرهای مستقل استفاده می‌شود، متغیرهای مستقل بایستی مستقل از هم بوده و با یکدیگر همبستگی نداشته باشند، چرا که باعث ایجاد هم‌خطی در رگرسیون می‌گردد. در صورت وجود هم‌خطی قوی، نتایج آزمون t قابل اعتماد نیست چرا که ضرایب رگرسیون ممکن است درست برآورد نشده باشند و یا ضریب بدست آمده اثر خالص آن متغیر نباشد. تشخیص هم‌خطی متغیرها در آمار و یادگیری ماشین به کمک پارامتر Variation Inflation Factor (VIF) صورت می‌گیرد که توضیح ریاضی این پارامتر خارج از این بحث است ولی $VIF > 10$ نشان‌دهنده هم‌خطی قوی در رگرسیون است. VIF تمامی متغیرهای مستقل در مدل رگرسیون خطی بر اساس نتایج آزمون t در جدول ۴ نشان داده شده است که همه کوچک‌تر از ۱۰ بوده و مدل از نظر هم‌خطی مشکلی نداشت.

ضرایب متغیرها تفاوت معناداری با صفر دارند. جهت ارزیابی فرض اول رگرسیون یعنی نرمال بودن توزیع خطاها، آزمون‌های آماری Moment شامل آزمون Jarque-Bera و آزمون Anscombe-Glynn انجام و نرمال بودن توزیع خطاها تأیید شد. جهت ارزیابی فرض دوم یعنی همسانی توزیع واریانس خطاها یا عدم Heteroscedasticity نمودار خطاها یا $(y - \hat{y})$ در برابر \hat{y} رسم می‌شود. انتظار این است که مقادیر $(y - \hat{y})$ حول صفر توزیع شده باشند و واریانس ثابتی داشته باشند. نمودار خطاها یا Residuals $(y - \hat{y})$ در برابر \hat{y} در شکل ۷ الف نشان‌دهنده کمی ناهمسانی در توزیع واریانس خطاهاست، بنابراین فرض دوم رگرسیون خطی نقض شده و مقادیر p-value آزمون‌های آماری، ضرایب رگرسیون و بازه‌های اطمینان رگرسیون خطی نامعتبر است و بایستی از دیگر مدل‌های رگرسیون خطی که الزام مفروضات فوق را ندارند و یا از مدل‌های رگرسیون غیرخطی برای پیش‌بینی استفاده نمود.

کیفیت مدل از نظر داده‌های نامتعارف و تأثیرگذار بر مدل نیز از دو روش Leverage و Cook's distance نیز بررسی شد که به طور خلاصه این روش‌ها تأثیر تک‌تک مشاهدات بر روی رگرسیون را ارزیابی می‌کند، به‌طوری‌که Cook's distance با فاصله بزرگ‌تر از یک نشان می‌دهد که آن مشاهده اثر زیادی روی مدل رگرسیون دارد. اصولاً وجود یا



شکل ۷. الف) نمودار خطاها یا Residuals در برابر \hat{y} ، ب) نمودار خطاها یا Residuals در برابر Leverage

جدول ۴. مقادیر VIF متغیرهای مستقل در مدل رگرسیون خطی

	α	UCS	ROD	GSI	Fn
VIF	1.62	1.83	1.75	1.73	1.54

MAE یا میانگین قدر مطلق خطاها در مدل Lm-t test برابر ۰/۱۹ بدست آمد، این بدان معناست که رابطه فوق به طور میانگین با ۰/۱۹ mm خطا در هر دور گردش کله حفار، می تواند مقدار نفوذ کله حفار را پیش بینی کند. جهت کاهش خطای پیش بینی در رگرسیون خطی و حل مشکل فرضیات رگرسیون، رویکردهای جدیدتری مثل رگرسیون گام به گام (Stepwise Regression) توسعه یافته اند، که خود شامل رویکردهای متنوعی مثل الگوریتم های Best subset reg.

Forward reg. و Backward reg. می شوند. ($R^2 = 0.86$)

در این الگوریتم ها وقتی هدف پیش بینی است، تا حد امکان مدل های رگرسیون ساخته شده و بهترین رگرسیون انتخاب می شود. در این رویکرد نیازی به برقرار بودن فرضیات رگرسیون برای معتبر بودن نتایج آزمون t نیست و فرض عدم هم خطی متغیرها هم در رگرسیون وجود ندارد. بهترین رگرسیون از طریق روش های آماری مثل adjusted-R^2 یا از روش یادگیری ماشین K-fold Cross Validation انتخاب می گردد. در روش Best subset regression تا حد امکان مدل های رگرسیون ساخته شده و بهترین رگرسیون بر اساس R^2 انتخاب می شود. بر خلاف روش Best subset reg. که در ساخت مدل توجهی به نتایج رگرسیون های قبلی ندارد، در دو الگوریتم دیگر Forward و Backward انتخاب متغیرها در هر رگرسیون با توجه به متغیرهای بهترین رگرسیون مرحله قبلی صورت می گیرد (Beckett et al., 2017).

در ادامه هر سه الگوریتم فوق روی داده های Train آموزش داده شد و بهترین مدل بر اساس adjusted-R^2 مدل چهار متغیره با رابطه (۸) است که در شکل ۸ الف نشان داده شده است و بهترین مدل بر اساس روش K-fold Cross Validation مدل دو متغیره با رابطه (۹) است که در شکل ۸ ب نشان داده شده است.

حداقل ۵ بار مدل رگرسیون کلاسیک بر روی داده ها با نمونه گیری تصادفی و تقسیم داده ها به دو بخش Train و Test اجرا و در همه مدل ها فقط دو پارامتر UCS و Fn متمایز شد. بنابراین مدل Linear model-t test براساس متغیرهای متمایز شده ی آزمون t ساخته شده و روی داده های Train آموزش داده شد. رابطه (۳) رابطه بدست آمده با R^2 در حدود ۰/۸۶ است.

$$PR = 1.84 - 0.035UCS + 0.035Fn \quad (3)$$

در این رابطه PR بر حسب mm/rev، UCS بر حسب MPa و Fn بر حسب kN می باشد. حال مدل را در معرض داده های Test قرار داده و برای اینکه دید بهتری نسبت به عملکرد مدل پیدا کنیم، توزیع خطاها را بر اساس پارامترهای ارزیابی مدل شامل Mean Absolute Error (MAE)، Mean Absolute Percentage Error (MAPE) بر حسب درصد، Squared Error (MSE) و Root Mean Squared Error (RMSE) می سنجیم. روابط (۴) تا (۷) روابط ریاضی این پارامترهاست. در این روابط y_i متغیر وابسته اندازه گیری شده، \hat{y}_i متغیر وابسته پیش بینی شده از طریق مدل و n تعداد مشاهدات است. مقادیر پارامترهای ارزیابی مدل برای هر مدل ساخته شده در هر مرحله در جدول ۶ آمده است. همچنین برای درک بهتر توزیع خطای مدل، آنالیز آماری قدرمطلق خطاها یعنی $|y_i - \hat{y}_i|$ برای هر مدل در جدول ۷ آمده است.

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n} \quad (4)$$

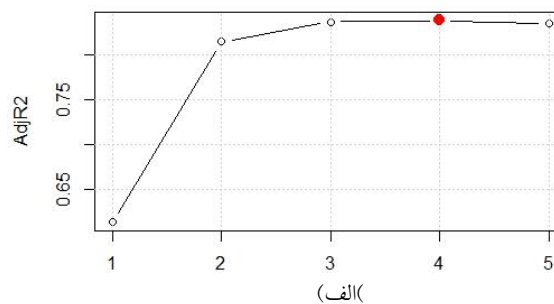
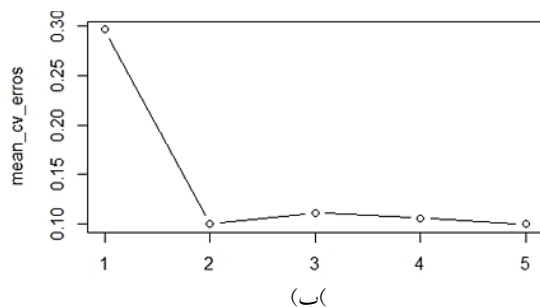
$$MAPE = \frac{1}{n} \sum \frac{|y_i - \hat{y}_i|}{y_i} * 100 \quad (5)$$

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

$$PR = 1.94 - 0.027UCS + 0.036Fn - 0.006RQD - 0.0085GSI \quad (R^2 = 0.86) \quad (8)$$

$$PR = 1.58 - 0.033UCS + 0.036Fn \quad (R^2 = 0.83) \quad (9)$$



شکل ۸. نمودار تعداد متغیرها در برابر الف) adjusted-R²، ب) K-fold CV در رگرسیون گام‌به‌گام

این پنالتی در Ridge Regression به صورت رابطه (۱۰) و در LASSO به صورت رابطه (۱۱) مینیمم می‌گردد (Hastie et al., 2009)

$$\min : SSE + \lambda \sum_{j=1}^m (\hat{\beta}_m)^2 \quad \lambda \geq 0 \quad (10)$$

$$\min : SSE + \lambda \sum_{j=1}^m |\hat{\beta}_m| \quad \lambda \geq 0 \quad (11)$$

تفاوت در دو رابطه فوق باعث تفاوت‌هایی در خروجی مدل می‌شود به طوری که در Ridge Reg. با بزرگ شدن λ ، $\hat{\beta}$ به سمت صفر میل می‌کند تا SSE مینیمم گردد ولی هیچ‌وقت صفر نمی‌شود ولی در LASSO، $\hat{\beta}$ می‌تواند صفر شود. نهایتاً در Ridge Reg. تمام متغیرها در مدل خواهند ماند حتی با ضرایب خیلی کوچک ولی در LASSO برخی از متغیرها با $\hat{\beta} = 0$ حذف می‌شوند. از این جهت الگوریتم LASSO رویکرد مدرن‌تری نسبت به Stepwise Reg. برای انتخاب متغیرهای مؤثر در رگرسیون است و زمانی که تنها برخی متغیرها برای پیش‌بینی متغیر وابسته مفید هستند، الگوریتم LASSO بهتر از Ridge Reg. است. در ادامه هر دو الگوریتم Ridge Reg. و LASSO روی داده‌ها آموزش داده شد. مدل Ridge Reg. با حضور تمامی متغیرهای مستقل با رابطه (۱۲) نشان می‌دهد که با بزرگ شدن λ ضرایب رگرسیون به سمت صفر میل می‌کند تا SSE به حداقل برسد ولی هیچ‌وقت صفر نمی‌شود. مدل LASSO نیز مدل چهار متغیره با رابطه (۱۳) می‌باشد.

به دلیل محدود بودن تعداد داده‌ها در این مطالعه (۴۱ دسته داده) مدل به دست آمده از هر سه الگوریتم Best subset reg. Forward reg. و Backward reg. یکسان بود و بهترین مدل با کمترین خطا بر اساس جدول ارزیابی مدل‌ها (جدول ۶) مدل دو متغیره بر اساس روش K-fold Cross Validation با MAE برابر ۰/۱۵ می‌باشد. همان‌طور که در جدول ۶ نشان داده شده است، R² در مدل‌های اخیر بهبود چندانی پیدا نکرد، پس می‌توان به این نتیجه رسید که با روش‌های رگرسیون خطی یا گام‌به‌گام بیش از این نمی‌توان مدل را بهبود بخشید. تا این مرحله، اولین رویکردهایی که در یادگیری ماشین برای پیش‌بینی متغیرهای پیوسته یا رگرسیون به کار می‌رود، پیاده‌سازی شد ولی از این پس الگوریتم‌های مدرن‌تر از جمله الگوریتم‌های قاعده‌سازی یا Regularization استفاده می‌شود. این رویکرد هم شبیه رگرسیون خطی است با این تفاوت که به عمد به مدل سوگیری یا Bias اضافه می‌کند تا مدل به بهترین شکل روی داده‌های Train آموزش ببیند، به امید این که مدل روی داده‌هایی که هنوز ندیده است از جمله داده‌های Test عملکرد بهتری داشته باشد. به این ترتیب با برآزش مدلی که روی داده‌های Train سوگیری دارد، واریانس را در داده‌های Test کم می‌کنیم تا به قدرت پیش‌بینی بهتری برسیم. این رویکرد به دو روش Ridge Regression و LASSO انجام می‌شود که در آن همانند روش OLS هدف کمینه کردن مجموع مجذور خطاهاست به علاوه یک پنالتی،

$$PR = 2.18 - 0.025UCS + 0.033Fn - 0.0056RQD - 0.0086GSI - 0.001\alpha \quad (R^2 = 0.86) \quad (12)$$

$$PR = 1.93 - 0.026UCS + 0.034Fn - 0.004RQD - 0.007GSI \quad (R^2 = 0.86) \quad (13)$$

مدل رگرسیون خطی با کمترین خطا، مدل Stepwise Regression با رویکرد K-fold Cross Validation با میانگین قدرمطلق خطاها (MAE) برابر ۰/۱۵ و پس از آن مدل LASSO با میانگین قدرمطلق خطاها (MAE) برابر ۰/۱۷ می‌باشد. آنچه مهم است، کارکرد مدل در موارد عملی است، به طور مثال مدل رگرسیون خطی، مدل ساده‌ای است و در پیاده‌سازی، به‌روز رسانی، ارتباط گرفتن با کاربر و غیره بسیار ساده است. بنابراین اگر بتوان مسئله را با رویکرد رگرسیون خطی حل کرد، نباید مدل را پیچیده‌تر نمود ولی اگر محدوده خطای مدل برای کاربر قابل قبول نباشد و یا R^2 بالاتری مدنظر باشد، باید مدل‌های پیچیده‌تر غیرخطی را پیاده‌سازی کرد.

سپس مدل‌ها را در معرض داده‌های Test قرار داده، پارامترهای ارزیابی مدل‌ها محاسبه و در جدول ۶ آمده است. همچنین برای درک بهتر توزیع خطای مدل، آنالیز آماری قدرمطلق خطاها یعنی $|y_i - \hat{y}_i|$ در جدول ۷ آمده است. توجه به این نکته ضروری است که رویکردهای Ridge Regression و LASSO هنوز در فضای خطی هستند و متغیر وابسته بر اساس ترکیب متغیرها، به صورت خطی مدل می‌شوند.

تمامی مدل‌های رگرسیون خطی ساخته شده بر اساس الگوریتم‌های مختلف در جدول ۵ خلاصه شده است. محدوده بازه اطمینان به دست آمده برای ضرایب هر متغیر مستقل در مدل‌ها نیز در آن ذکر شده است. با توجه به جدول ۶ بهترین

جدول ۵. مدل‌های رگرسیون خطی ساخته شده بر روی داده‌های تونل SEL6 در شرایط آسیاب‌کاری

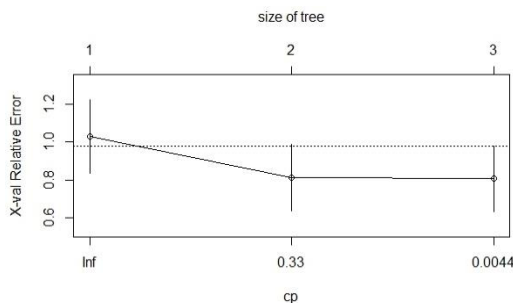
Model	Equation	Confidence Intervals for model parameters (up to 95%)
Lm-t test	$PR = 1.84 - 0.035UCS + 0.035Fn$	Intercept: -0.55 - 3.9 UCS: -0.041 - -0.025 Fn: 0.022 - 0.048
Best subset-Adj- R^2 Forward-Adj- R^2 Backward-Adj- R^2	$PR = 1.94 - 0.027UCS + 0.036Fn - 0.006RQD - 0.0085GSI$	Intercept: -0.18 - 4.06 UCS: -0.04 - -0.014 Fn: 0.024 - 0.047 GSI: -0.02 - 0.003 RQD: -0.016 - 0.005
Best subset-CV Forward-CV Backward-CV	$PR = 1.58 - 0.033UCS + 0.036Fn$	Intercept: -0.66 - 3.81 UCS: -0.045 - -0.021 Fn: 0.02 - 0.048
Ridge reg.	$PR = 2.18 - 0.025UCS + 0.033Fn - 0.0056RQD - 0.0086GSI - 0.001\alpha$	-
LASSO	$PR = 1.93 - 0.026UCS + 0.034Fn - 0.004RQD - 0.007GSI$	-

Boosting استفاده کرد. همانند الگوریتم‌های رگرسیون خطی در بخش قبلی، در این بخش از مطالعه نیز الگوریتم‌های رگرسیون غیرخطی نیز با زبان برنامه‌نویسی R پیاده‌سازی شده‌اند. یکی از ساده‌ترین الگوریتم‌های رگرسیون غیرخطی، الگوریتم درخت تصمیم است که ساختاری درخت مانند دارد و بر حسب اینکه متغیر وابسته که می‌خواهیم پیش‌بینی کنیم،

۴-۲. رگرسیون غیرخطی

وقتی احتمال غیرخطی بودن رابطه بین متغیرها وجود دارد و یا رابطه بین متغیر وابسته و متغیرهای مستقل شناسایی نشده باشند، می‌توان از الگوریتم‌های رگرسیون غیرخطی مثل درخت تصمیم (Decision Tree) و یا رویکردهای یادگیری جمعی (Ensemble Learning) مثل الگوریتم‌های Bagging و

نشان می‌دهد، بالاترین Terminal node، F_n است که کمترین خطای SSE را در داده‌ها داشته و بنابراین در بالاترین node قرار می‌گیرد. مقدار نفوذ PR پیش‌بینی شده برای مشاهدات با $F_n > 126 \text{ kN}$ برابر (3.7 mm/rev) و مشاهدات با $F_n < 126 \text{ kN}$ باید از نظر UCS بررسی شوند. اگر $UCS \geq 100$ باشد، PR برابر (1.9 mm/rev) و اگر $UCS < 100$ باشد، برابر (2.6 mm/rev) پیش‌بینی می‌شود. حال الگوریتم را در معرض داده‌های Test قرار داده، تا عملکرد مدل ارزیابی شود. با توجه به جدول ۶ و بررسی پارامترهای ارزیابی مدل مشاهده می‌شود که R^2 مدل کاهش یافته و محدوده خطاهای پیش‌بینی نسبت به الگوریتم‌های قبلی بیش از دو برابر شده است، پس الگوریتم درخت تصمیم، مدل مناسبی برای پیش‌بینی PR نیست و الگوریتم‌های رگرسیون خطی مناسب‌ترند. یکی از دلایل می‌تواند کم بودن تعداد داده‌های مورد مطالعه باشد، چراکه این الگوریتم برای مدل‌سازی روی داده‌های بزرگ با متغیرهای پیچیده و روابط غیرخطی مناسب است.



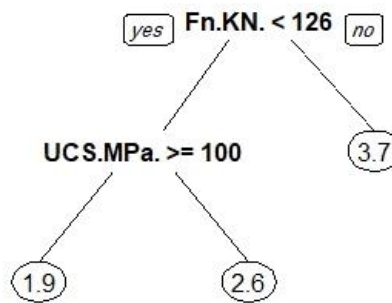
شکل ۹. درخت رگرسیون حاصل از آموزش الگوریتم درخت تصمیم برای پیش‌بینی PR در شرایط آسیاب‌کاری

روش‌های Black Box شناخته می‌شوند، چراکه قوانین یا الگوهای ضمنی ساخت مدل به راحتی قابل تفسیر و توضیح نیست. بنابراین مزیت اصلی استفاده از الگوریتم درخت تصمیم این است که ساختار درختی را می‌توان به عنوان یک روش White Box ارائه، به راحتی تفسیر و تعریف کرد (Salimi, 2021).

الگوریتم‌های Bagging و Random Forest تا حد زیادی

پیوسته (Continuous) یا رسته‌ای (Categorical) باشد، بصورت Regression Tree یا Classification Tree ساخته می‌شود. در این ساختار درختی در رأس هر شاخه یا node یک شرط وجود دارد که اگر این شرط برقرار باشد، وارد شاخه سمت چپ می‌شویم و یک عدد برای پیش‌بینی اعلام می‌شود. اگر شرط برقرار نباشد، وارد شاخه سمت راست می‌شویم، در این مرحله شرط بعدی بررسی می‌شود و دوباره وارد شاخه سمت راست یا چپ می‌شود. به این عملکرد در یادگیری ماشین، جداسازی دوگانه (Binary) گفته می‌شود. به این ترتیب در رویکرد درخت تصمیم، فضای داده‌ها گسسته می‌شود و برای هر بازه‌ای از داده‌ها، یک عددی را پیش‌بینی می‌کند که این عدد میانگین متغیر وابسته در آن بازه است. درخت تصمیم حاصل آموزش الگوریتم درخت تصمیم بر روی داده‌های Train می‌باشد، این الگوریتم زمانی متوقف می‌گردد که تعداد مشاهدات در حداقل یک برگ، از مقدار مشخصی کمتر گردد (Breiman et al., 1984).

همان‌طور که مدل درخت رگرسیون حاصل از آموزش الگوریتم درخت تصمیم بر روی داده‌های Train در شکل ۹



الگوریتم درخت تصمیم مزایا و معایبی دارد؛ مزیت آن توانایی مدل‌سازی مسائل پیچیده و غیرخطی است، همچنین مدل ساده‌ای است و عملاً با استفاده از میانگین‌گیری کار می‌کند و برخلاف مدل‌های Black Box به راحتی برای کاربران توضیح داده می‌شود. به عنوان مثال برخی تکنیک‌های هوش مصنوعی مانند ANN یا SVR به عنوان تکنیک‌های قدرتمندی در تخمین پارامترهای عملکرد TBM هستند، اما به عنوان

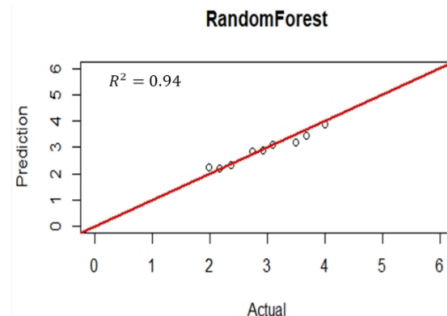
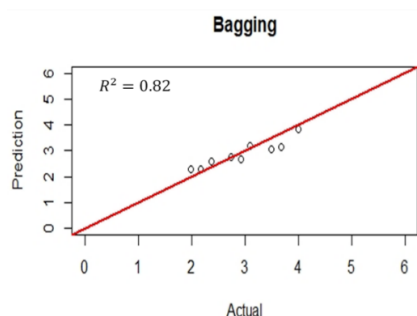
بیشتری داده تا پیش‌بینی صورت گیرد. منطق تعیین این شاخص‌ها به این صورت است که چون در Random Forest برخی متغیرها به صورت تصادفی در مدل نیستند، می‌توان فهمید که عدم حضور یک متغیر در مدل، چقدر خطای MSE مدل را زیاد می‌کند. خطای پیش‌بینی در داده‌های Out Of Bag (OOB) محاسبه می‌شوند، چراکه داده‌های OOB تاکنون در معرض الگوریتم قرار نگرفته و برای محاسبه خطای پیش‌بینی بسیار مناسب‌اند (Hastie et al., 2021). این شاخص‌ها بنام شاخص اهمیت متغیر یا Variable Importance با دو پارامتر افزایش خطا MSE و کاهش ناخالصی گره (Node Impurity)، متغیرهای مهم‌تر در الگوریتم را متمایز می‌کنند و متغیرهایی که وجود آن‌ها تأثیر چندانی در کاهش خطای مدل ندارند و مقدار Variable Importance آن صفر یا نزدیک صفر است، می‌توانند حذف شوند. مقادیر Variable Importance مدل، حاصل از آموزش الگوریتم Random Forest بر روی داده‌های Train نشان داد که تنها F_n و UCS شاخص Variable Importance بالایی دارند. هر دو الگوریتم Bagging و Random Forest بر اساس هابیر پارامترهای بهینه بر روی داده‌های Train پیاده‌سازی شدند و سپس مدل‌ها در معرض داده‌های Test قرار گرفته تا عملکرد مدل ارزیابی گردد. با توجه به جدول ۶ و بررسی پارامترهای ارزیابی مدل، مشاهده می‌شود که R^2 مدل، هم در داده‌های Train و هم در داده‌های Test افزایش یافته و در مدل Random Forest به ۰/۹۴ رسیده که ضریب بسیار خوبی است. میانگین قدرمطلق خطاها (MAE) برابر ۰/۱۳ شده و کاهش محسوسی یافته، در نتیجه Random Forest مدل مناسبی برای پیش‌بینی PR است و تنها نقطه ضعف مدل این است که به دلیل پیچیدگی الگوریتم، روابط بین متغیرها قابل تفسیر نیست و باید مدل ساخته شده به نحوی قابل استفاده برای عموم مهندسين تونل و ژئوتکنیک درآید. مقادیر PR واقعی داده‌های Test در برابر مقادیر پیش‌بینی شده آن توسط مدل‌های Bagging و Random Forest در شکل ۱۰ نشان داده شده است.

به هم شبیه هستند، به طوری که این روش‌ها ترکیب الگوریتم درخت تصمیم و روش آماری نمونه‌گیری تصادفی با جای‌گذاری به نام Bootstrap می‌باشند. این‌که الگوریتم درخت تصمیم یک روش وابسته به نمونه (Sample) است و با تغییر نمونه، خروجی مدل تغییرات اساسی پیدا می‌کند، یک ایراد مهم در این الگوریتم است. برای بهبود این ضعف الگوریتم‌های Bagging و Random Forest توسعه یافته‌اند که در این الگوریتم‌ها صدها بار به روش Bootstrap نمونه‌گیری صورت می‌گیرد و روی این نمونه‌ها، صدها درخت رگرسیون ساخته می‌شود. در این حالت پیش‌بینی متغیر وابسته به جای یک درخت براساس صدها درخت صورت می‌گیرد و میانگین پیش‌بینی این صدها درخت به عنوان پیش‌بینی الگوریتم بیان می‌شود (Hastie et al., 2021). تفاوت این دو الگوریتم در این است که در رویکرد Bagging الگوریتم اجازه دارد که از تمام متغیرهای مستقل موجود برای ساخت مدل استفاده کند، ولی در الگوریتم Random Forest در معرض تمام متغیرهای مستقل موجود نیست و هر بار الگوریتم در معرض زیرمجموعه‌ای تصادفی از متغیرهای مستقل قرار می‌گیرد و درخت‌ها ساخته می‌شوند. این موضوع به جلوگیری از مسئله برازش بیش‌ازحد که در الگوریتم درخت تصمیم معمولاً باعث پایین آمدن عملکرد مدل می‌شود، کمک می‌کند (Boehmke & Greenwell, 2020).

الگوریتم Random Forest الگوریتم بسیار قدرتمندی است ولی نکته منفی آنست که به علت پیچیدگی الگوریتم، خروجی مدل به صورت Black Box است. همان‌طور که گفته شد در الگوریتم درخت تصمیم به راحتی منطق و چگونگی شکل‌گیری درخت و بازه‌های هر متغیر مستقل قابل‌تعریف است و یا مثلاً در رگرسیون خطی ضرایب هر متغیر تعیین شده و از طریق بزرگی ضریب هر متغیر می‌توان به اهمیت آن متغیر در خط رگرسیون پی برد. ولی الگوریتم‌های Bagging و Random Forest پیچیدگی زیادی دارند. برای شفاف شدن نتایج مدل این الگوریتم‌ها شاخص‌هایی معرفی شده‌اند که از طریق آن می‌توان دریافت که مدل به کدام متغیرها وزن

در الگوریتم‌های دیگری تحت عنوان تقویت گرادیان یا Gradient boost در مرحله اول یک مدل ساده مثل میانگین‌گیری روی داده‌ها پیاده‌سازی می‌شوند و روی داده‌ها کار می‌کند و یک پیش‌بینی اعلام می‌کند. واضح است که ساده‌ترین پیش‌بینی متغیر وابسته در یادگیری نظارت‌شده، میانگین y است. این پیش‌بینی طبیعتاً خطا دارد. در مرحله بعد Learner بعدی وارد عمل شده و سعی می‌کند مدل قبلی را بهبود بخشد و به صورت سری این کار ادامه می‌یابد تا بهترین پیش‌بینی روی داده‌ها صورت گیرد. این الگوریتم همانند Boost Reg. (2000) و XGBoost (2014) می‌باشند (Friedman, 1999).

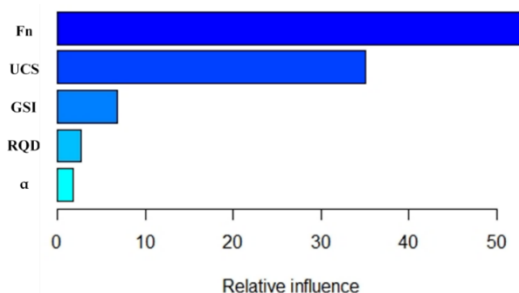
در الگوریتم‌های دیگری تحت عنوان تقویت گرادیان یا Gradient boost در مرحله اول یک مدل ساده مثل میانگین‌گیری روی داده‌ها پیاده‌سازی می‌شوند و روی داده‌ها کار می‌کند و یک پیش‌بینی اعلام می‌کند. واضح است که ساده‌ترین پیش‌بینی متغیر وابسته در یادگیری نظارت‌شده، میانگین y است. این پیش‌بینی طبیعتاً خطا دارد. در مرحله بعد Learner بعدی وارد عمل شده و سعی می‌کند مدل قبلی را بهبود بخشد و به صورت سری این کار ادامه می‌یابد تا بهترین پیش‌بینی روی داده‌ها صورت گیرد. این الگوریتم همانند Boost Reg. (2000) و XGBoost (2014) می‌باشند (Friedman, 1999).



شکل ۱۰. مقادیر PR واقعی در برابر مقادیر پیش‌بینی شده توسط مدل‌های Bagging و RandomForest

هر دو الگوریتم Gradient Boost Reg. و XGBoost بر اساس هایپرپارامترهای بهینه بر روی داده‌های Train اجرا و سپس مدل‌ها در معرض داده‌های Test قرار گرفته تا عملکرد مدل ارزیابی شود. با توجه به جدول ۶ و بررسی پارامترهای ارزیابی مدل مشاهده می‌شود که R^2 مدل، در داده‌های Train افزایش و در داده‌های Test کاهش یافته است که ممکن است به دلیل برازش بیش‌ازحد روی داده‌های Train باشد. میانگین قدرمطلق خطاها (MAE) در Gradient Boost برابر ۰/۲ و در XGBoost برابر ۰/۲۴ است و خطاها نسبت به مدل Random Forest افزایش یافته، بنابراین می‌توان نتیجه گرفت Random Forest مدل مناسب‌تری برای پیش‌بینی PR است. در این تحقیق ۱۰ مدل الگوریتم رگرسیون مختلف، روی داده‌های مورد مطالعه پیاده‌سازی شده که هر کدام رویکردهای متفاوتی در ساخت مدل داشتند. مدلهایی که بایستی Tuning می‌شدند، با روش K-fold Cross Validation هایپرپارامترهای

در انواع الگوریتم‌های تقویت گرادیان پارامتر Relative Influence شبیه شاخص Variable Importance در الگوریتم Random Forest، نشان می‌دهد که اهمیت هر پارامتر در رگرسیون چقدر است و نبود هر کدام از این متغیرها، چقدر باعث افزایش خطا در مدل می‌شود. مقادیر Relative Influence مدل حاصل از آموزش الگوریتم XGBoost بر روی داده‌های Train در شکل ۱۱ نشان داده شده است که همانند الگوریتم‌های قبلی، تنها F_n و UCS اهمیت بالایی دارند.



شکل ۱۱. مقادیر Relative Influence مدل XGBoost

غیرخطی در داده‌ها وجود داشته باشد، می‌توان از ساده‌ترین مدل‌های غیرخطی مثل درخت تصمیم استفاده کرد. در داده‌های مورد مطالعه به دلیل محدود بودن تعداد مشاهدات، رویکرد درخت تصمیم ساده خطای مدل را افزایش داد ولی رویکردهای مدرن‌تر مثل Bagging و Random Forest عملکرد بهتری داشت. الگوریتم‌های Boosting هم روی داده‌ها پیاده‌سازی شد و به دلیل این‌که بهبودی در عملکرد مدل حاصل نشد، فقط از دو الگوریتم Gradient Boost و XGBoost استفاده شد. آنچه مهم است این است که نتایج مدل‌های ارائه شده بایستی Robust بوده و با تغییر داده‌های Train و Test و نمونه‌گیری مجدد، با تغییر بازه اندکی، همین نتایج بدست آید.

بهینه انتخاب و مدل نهایی ساخته شد و سپس مدل روی داده‌های Test پیاده‌سازی و عملکرد مدل بدست آمده است. همان‌طور که قبلاً نیز ذکر شد، پارامترهای ارزیابی مدل الگوریتم‌های مختلف در جدول ۶ و آنالیز آماری قدرمطلق خطاها $|y_i - \hat{y}_i|$ برای هر مدل در جدول ۷ ارائه شده است. باید توجه داشت که فرآیندهای یادگیری ماشین همواره تابع فرایندهای تصادفی است و این اعداد مطلق نیستند. عموماً مدل‌سازی با مدل‌های خطی بر اساس نتایج آزمون t شروع شده و سپس به کمک الگوریتم‌های رگرسیون گام‌به‌گام سعی در بهبود عملکرد مدل داشته و از آنجاکه با این الگوریتم‌ها بهبود چندانی در مدل حاصل نشد، سعی شد با استفاده از رویکردهای Regularization، همچنان در فضای رگرسیون خطی عملکرد مدل بهبود یابد. اگر احتمال وجود روابط

جدول ۶. مقایسه پارامترهای ارزیابی مدل در الگوریتم‌های مختلف

Model	R ²		MAE	MAPE (%)	MSE	RMSE
	Train	Test				
Lm-t test	0.86	0.91	0.19	6.5	0.04	0.2
Best subset-Adj-R ²						
Forward-Adj-R ²	0.86	0.92	0.18	6.3	0.038	0.2
Backward-Adj-R ²						
Best subset-CV						
Forward-CV	0.83	0.93	0.15	5.2	0.032	0.18
Backward-CV						
Ridge reg.	0.86	0.92	0.18	6.3	0.036	0.19
LASSO	0.86	0.93	0.17	6.0	0.033	0.18
Regression Tree	0.57	0.41	0.44	15	0.27	0.52
Bagging	0.94	0.82	0.24	8.2	0.081	0.28
Random Forest	0.93	0.94	0.13	4.5	0.027	0.16
Gradient Boost	0.93	0.88	0.20	7.1	0.054	0.23
XGBoost	0.99	0.77	0.24	8.5	0.107	0.32

جدول ۷. آنالیز آماری قدرمطلق خطاها $|y_i - \hat{y}_i|$ برای هر مدل

Model	Mean of MAE	Median of MAE	SD of MAE	IQR of MAE	Min of MAE	Max of MAE
Lm-t test	0.187	0.143	0.085	0.128	0.106	0.336
Best subset-Adj-R ²						
Forward-Adj-R ²	0.177	0.209	0.091	0.145	0.053	0.296
Backward-Adj-R ²						
Best subset-CV						
Forward-CV	0.145	0.119	0.111	0.149	0.011	0.320
Backward-CV						
Ridge reg.	0.177	0.202	0.070	0.085	0.070	0.288
LASSO	0.167	0.189	0.079	0.089	0.067	0.309
Regression Tree	0.436	0.408	0.305	0.533	0.015	0.902
Bagging	0.241	0.233	0.162	0.175	0.022	0.525
Random Forest	0.129	0.110	0.109	0.195	0.010	0.306
Gradient Boost	0.207	0.185	0.112	0.090	0.057	0.453
XGBoost	0.239	0.102	0.237	0.308	0.003	0.733

۵. نتیجه‌گیری

رگرسیون به یک تکنیک آماری در یادگیری ماشین و داده‌کاوی گفته می‌شود که به کمک آن رابطه بین یک یا چند متغیر، مدل‌سازی می‌شود. شیوه‌های مختلفی برای برآزش خط یا منحنی روی داده‌ها وجود دارد که در این مطالعه ۱۰ مدل مختلف رگرسیون پیاده‌سازی شد. مدل‌سازی با مدل‌های خطی بر اساس نتایج آزمون t شروع و سپس به کمک الگوریتم‌های Stepwise Regression سعی در بهبود عملکرد مدل شد. پس از آن با استفاده از رویکردهای Regularization همچنان در فضای رگرسیون خطی عملکرد مدل بهبود یافت و بهترین مدل رگرسیون خطی با کمترین خطا را، مدل Stepwise Regression با رویکرد K-fold Cross Validation با میانگین قدرمطلق خطاها (MAE) برابر ۰/۱۵ داشت. در مرحله بعد مدل غیرخطی درخت تصمیم پیاده‌سازی شد که به دلیل محدود بودن تعداد مشاهدات، خطای مدل افزایش یافت، ولی رویکردهای مدرن‌تر Bagging و Random Forest با انتخاب بهینه هایپرپارامترها، عملکرد بهتری روی داده‌ها داشت و مدل Random Forest با R^2 بالا و میانگین قدرمطلق خطاها (MAE) برابر ۰/۱۳ کمترین خطا را در بین مدل‌ها داشت. در رگرسیون خطی از یک Learner برای پیش‌بینی استفاده می‌شود، ولی در رویکردهای Bagging و Random

Forest عملاً از تعداد زیادی Learner استفاده می‌شود و مدل‌ها Train شده و از میانگین پیش‌بینی‌ها به‌عنوان پیش‌بینی مدل استفاده می‌شود. پس می‌توان نتیجه گرفت که رویکردهای یادگیری جمعی یا Ensemble Learning نتایج بهتری دارند و از بین آن‌ها نیز Random Forest بهترین عملکرد را داشت. یکی از مشکلات مرتبط با استفاده از تکنیک‌های AI، استفاده محدود آن برای مهندسی تونل‌سازی و ژئوتکنیک است. استفاده از یک مدل Random Forest توسط یک مهندس تونل مستلزم آن است که شخص به علم داده و رویکردهای AI آشنا باشد، پس باید مدل ساخته شده به‌نحوی قابل‌استفاده برای عموم مهندسی تونل و ژئوتکنیک درآید.

منبع اخذ داده‌ها

منبع داده‌های مورد استفاده در این مطالعه، بانک اطلاعاتی تهیه شده توسط یکی از نویسندگان همکار این پژوهش، جناب آقای دکتر حسن‌پور می‌باشد. آقای دکتر حسن‌پور مجری مطالعات زمین‌شناسی فاز یک و دو این پروژه بوده و حین ساخت تونل نیز همکاری نزدیکی با ارکان پروژه (کارفرما، مشاور و پیمانکار طرح) داشته‌اند.

منابع

- حسن‌پور، ج.، رستمی، ج.، خامه‌چیان، م.، ۱۴۰۲. زمین‌شناسی مهندسی و تونل‌سازی: ارزیابی عملکرد TBM در تونل‌های سنگی، انتشارات دانشگاه تهران، ۴۷۹ صفحه.
- شرکت مهندسی مشاور ره‌ساز طرح، ۱۳۹۵. گزارش مطالعات زمین‌شناسی مهندسی مسیر تونل SEL6.
- Armaghani, D. J., Mohamad, E. T., Narayanasamy, M. S., Narita, N., Yagiz, S., 2017. Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. *Tunnelling and Underground Space Technology*, 63, 29–43.
- Beckett, C., Eriksson, L., Johansson, E., & Wikström, C., 2017. Multivariate Data Analysis (MVDA). In *Pharmaceutical Quality by Design: A Practical Approach*.
- Boehmke, B., Greenwell, B., 2020. *Hands-On Machine Learning with R*. New York: Chapman and Hall/CRC, 488p.
- Breiman, L., Friedman, J., Stone, 1984. *Classification and Regression Trees*. Chapman and Hall/CRC.
- Bruland, A., 1998. *Hard rock tunnel boring*, Doctoral thesis, Norwegian University of Science and Technology, Trondheim
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

- Firouzei, Y., Hassanpour, J., Pourhashemi, S. M., 2019. Tunneling with a soft rock EPB machine in hard rock conditions, the experience of Tehran metro line 6 southern expansion sector. 4th International Conference on Tunnel Boring Machines in Difficult Grounds.
- Frenzel, C., Kasling, H., Thuro, K., 2008. Factors Influencing Disc Cutter Wear. *Geomechanics and Tunnelling: An International Journal*, 1(1): 55-60.
- Friedman, J. H., 1999. Stochastic Gradient Boosting. Technical report, Department of Statistics, Stanford University.
- Gehring, K., 2009. The influence of TBM design and machine features on performance and tool wear in rock. *Geomechanics and Tunnelling* 2(2), 140-155.
- Girmscheid, G., 2008. *Baubetrieb und Bauverfahren im Tunnelbau*. Ernst & Sohn, Berlin, Germany
- Grima, A., Bruines, M., Verhoef, P. N. W., 2000. Modeling tunnel boring machine performance by neuro-fuzzy methods. *Tunnelling and Underground Space Technology*, 15(3), 259-269.
- Hassanpour, J., Rostami, J., Zhao, J., 2011. A new hard rock TBM performance prediction model for project planning. *Tunnelling and Underground Space Technology: An International Journal*, 26: 595-603.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer New York: Springer Series in Statistics: 2nd Edition, 745p.
- Hastie, T., Tibshirani, R., James, G., & Witten, D., 2021. *An introduction to statistical learning* (2nd ed.). Springer Texts, 102, 618p.
- He, B., Armaghani, D. J., & Lai, S. H., 2022. A Short Overview of Soft Computing Techniques in Tunnel Construction. *The Open Construction & Building Technology Journal*, 16(1), 1-6.
- Koopialipour, M., Nikouei, S. S., Marto, A., Fahimifar, A., Jahed Armaghani, D., Mohamad, E. T., 2019. Predicting tunnel boring machine performance through a model based on the group method of data handling. *Bulletin of Engineering Geology and the Environment*, 78(5), 3799-3813.
- Pourhashemi, S. M., Ahangari, A., Hassanpour, J., Eftekhari, M., 2021. Evaluating the influence of engineering geological parameters on TBM performance during grinding process in limestone strata. *Bulletin of Engineering Geology and the Environment*, 80: 3023-3040.
- Robbins, R. J., 1970. Economic factors in tunnel boring. In: *Proceedings of the South African tunneling conference: the technology and potential of tunneling*, vol. 1., Johannesburg.
- Rostami, J., Ozdemir, L., 1993. A New Model for Performance Prediction of Hard Rock TBMs. *Proceedings of the Rapid Excavation and Tunneling Conference*, Boston.
- Salimi, A., 2021. Investigation and Evaluation of Rock Mass Characteristics for Development of New TBM Performance Prediction Model in Hard Rock Conditions. University Stuttgart.
- Salimi, A., Rostami, J., Moormann, C., & Delisio, A., 2016. Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs. *Tunnelling and Underground Space Technology*, 58, 236-246.
- Salimi, A., Rostami, J., Moormann, C., 2019. Application of rock mass classification systems for performance estimation of rock TBMs using regression tree and artificial intelligence algorithms. *Tunnelling and Underground Space Technology*, 92.
- Tarigh Azali, S., Ghafoori, M., Lashkaripour, GR., Hassanpour, J., 2013. Engineering geological investigations of mechanized tunneling in soft ground: a case study, East-West lot of line 7, Tehran Metro, Iran. *Engineering Geology*, 166,170-185.
- Villeneuve, M.C., 2017. Hard rock tunnel boring machine penetration test as an indicator of chipping process efficiency. *Journal of Rock Mechanics and Geotechnical Engineering*, 9 (4), 611-622.
- Yagiz, S., Karahan, H., 2011. Prediction of hard rock TBM penetration rate using particle swarm optimization. *International Journal of Rock Mechanics and Mining Sciences*, 48(3), 427-433.
- Zhou, J., Bejarbaneh, B. Y., 2019. Forecasting of TBM advance rate in hard rock condition based on artificial neural network and genetic programming techniques. *Bulletin of Engineering Geology and the Environment*.