

مقاله علمی پژوهشی

ارزیابی نفوذپذیری نسبی نفت و آب با استفاده از یادگیری ماشین: مطالعه موردی از جنوب غرب ایران

علی رنجبر^{۱*}؛ محمدرسول دهقانی^۲؛ مهدی مالکی^۳

۱- استادیار گروه مهندسی نفت، دانشکده مهندسی نفت، گاز و پتروشیمی، دانشگاه خلیج فارس

۲- دانشجوی دکتری مهندسی نفت، دانشکده مهندسی نفت، دانشگاه شهید باهنر

۳- فارغ التحصیل گروه مهندسی نفت، دانشکده مهندسی نفت، گاز و پتروشیمی، دانشگاه خلیج فارس

دریافت مقاله: ۱۴۰۴/۰۴/۲۱ پذیرش مقاله: ۱۴۰۴/۰۹/۰۱

شناسه دیجیتال (DOI): 10.22107/ggj.2025.559197.1260

چکیده

واژگان کلیدی

نفوذپذیری نسبی، *CatBoost*،

آنالیز مغزه، *Extra trees*،

تحلیل *SHAP*

نفوذپذیری نسبی یکی از پارامترهای پتروفیزیکی کلیدی است که رفتار جریان چندفازی در محیط متخلخل را کنترل کرده و تأثیر قابل توجهی بر دقت شبیه‌سازی مخزن، پیش‌بینی برداشت و برنامه‌ریزی ازدیاد برداشت دارد. روش‌های آزمایشگاهی تعیین این پارامتر دقیق ولی زمان‌بر و پرهزینه‌اند. این پژوهش با بهره‌گیری از تکنیک‌های یادگیری ماشین و داده‌های مغزه‌ای از یک مخزن در جنوب غرب ایران، به برآورد نفوذپذیری نسبی نفت و آب پرداخته است. شازده نمونه مغزه با هفت ویژگی ورودی برای آموزش چهار مدل *Extra Trees*، *K-Nearest Neighbors*، *Categorical Boosting* و *Extreme Gradient Boosting* استفاده شد. مدل‌ها با روش *Bayesian Hyperparameter Tuning* بهینه و با شاخص‌های R^2 ، $RMSE$ و MAE ارزیابی شدند. در برآورد نفوذپذیری نسبی آب، مدل *Extra Trees* بهترین عملکرد را با $R^2=0.9974$ ، $RMSE=0.0045$ و $MAE=0.0007$ نشان داد. مدل *KNN* نیز به‌ویژه در بازه نفوذپذیری ۰.۱ تا ۰.۲ عملکرد مطلوبی داشت. در برآورد نفوذپذیری نسبی نفت، مدل *KNN* دقیق‌ترین نتایج را با $R^2=0.9973$ ، $RMSE=0.0113$ و $MAE=0.0024$ ارائه کرد، در حالی که مدل *Extra Trees* در محدوده‌های بالای نفوذپذیری ضعیف‌تر عمل کرد. تحلیل حساسیت *SHAP* نشان داد اشباع آب مؤثرترین عامل در هر دو مدل است؛ نفوذپذیری نفت در اشباع آب اولیه برای مدل آب و ویسکوزیته نفت برای مدل نفت اهمیت بالایی دارند. نتایج نشان می‌دهد یادگیری ماشین روشی توانمند و کارآمد برای جایگزینی آزمون‌های آزمایشگاهی در برآورد نفوذپذیری نسبی است و انتخاب مدل به نوع سیال هدف وابسته بوده و می‌تواند موجب بهبود شبیه‌سازی، کاهش هزینه‌ها و مدیریت بهینه مخزن شود.

۱. پیش‌گفتار

نقشی اساسی دارد. نفوذپذیری نسبی به‌صورت نسبت نفوذپذیری مؤثر یک سیال در یک اشباع معین به نفوذپذیری مطلق محیط تعریف می‌شود [۴، ۵]. نفوذپذیری مطلق، نشان‌دهنده قابلیت انتقال یک سیال منفرد در محیط متخلخل است [۶-۸]. با این حال، در اغلب سامانه‌های طبیعی، چند فاز سیال مانند نفت، آب و گاز به‌طور هم‌زمان در فضای منفذی حضور دارند و با یکدیگر برهم‌کنش می‌کنند. وجود این فازهای چندگانه موجب کاهش

نفوذپذیری نسبی مفهومی اساسی در درک جریان چندفازی در محیط‌های متخلخل است و در حوزه‌هایی نظیر مهندسی نفت، هیدروژئولوژی و علم مواد اهمیت ویژه‌ای دارد. این پارامتر توانایی یک محیط متخلخل را در عبور یک فاز سیال در حضور فازهای نامخلوط دیگر توصیف می‌کند [۱-۳]. نفوذپذیری نسبی برای مدل‌سازی جریان سیالات در شبیه‌سازی مخزن، حرکت آب زیرزمینی و رفتار مواد مختلف

نظریه پراکندگی^{۱۰} یا همبستگی‌های تجربی توسعه یافته‌اند [۱۶].
[۱۷].

در سال‌های اخیر، چند عامل موجب گسترش استفاده از روش‌های یادگیری ماشین^{۱۱} در برآورد نفوذپذیری نسبی در مهندسی نفت و شبیه‌سازی مخزن شده است:

- زمان‌بر و پرهزینه بودن آزمایش‌ها: روش‌های سنتی آزمایشگاهی اغلب وقت‌گیر و پرهزینه‌اند [۱۸] در حالی که یادگیری ماشین جایگزینی سریع‌تر و اقتصادی‌تر ارائه می‌دهد [۱۹].

- محدودیت در به‌روزرسانی بلادرنگ^{۱۲}: روابط سنتی نفوذپذیری نسبی از نظر تعمیم‌پذیری محدود بوده و به‌سختی برای سامانه‌های مختلف به‌صورت بلادرنگ قابل تنظیم هستند، در حالی که مدل‌های یادگیری ماشین به‌سادگی با داده‌های جدید به‌روزرسانی می‌شوند [۲۰].

- پیچیدگی در هم‌مقیاس‌سازی^{۱۳}: روش‌های سنتی هم‌مقیاس‌سازی دو فازی پرهزینه و زمان‌بر هستند، به‌ویژه برای تحلیل عدم قطعیت در چندین شبیه‌سازی زمین‌شناسی. یادگیری ماشین می‌تواند با یادگیری از بخشی از نتایج شبیه‌سازی دقیق، پیش‌بینی سایر موارد را انجام دهد.

- چالش در سامانه‌های دوترشوندگی^{۱۴}: مدل‌های سنتی مبتنی بر ترشوندگی یکنواخت برای سنگ‌های نامتعارف با شبکه‌های منفذی دوترشوندگی مناسب نیستند، در حالی که یادگیری ماشین قادر به مدل‌سازی این روابط پیچیده است [۲۱].

- ادغام داده‌های چندبعدی: یادگیری ماشین توانایی ادغام داده‌های چندبعدی از منابع مختلف مانند لاگ چاه و ویژگی‌های لرزه‌ای را برای بهبود برآورد نفوذپذیری دارد [۲۰].

- روابط غیرخطی: روابط میان نفوذپذیری نسبی و پارامترهای مخزن معمولاً پیچیده و غیرخطی هستند که یادگیری ماشین قابلیت شناسایی و مدل‌سازی آن‌ها را دارد [۲۲].

- ناهمگنی: مخازن کربناته معمولاً دارای ناهمگنی بالایی

قابلیت جریان هر فاز می‌شود و نفوذپذیری نسبی میزان این کاهش را کمی‌سازی می‌کند [۹-۱۱].

نفوذپذیری نسبی تحت تأثیر عوامل متعددی از جمله اشباع سیال، ویژگی‌های ترشوندگی، کشش بین‌سطحی، ساختار منفذی و پدیده هیستریزیس^۱ قرار دارد. اشباع سیال نقش کلیدی دارد، زیرا نفوذپذیری نسبی هر فاز معمولاً با افزایش اشباع آن در محیط متخلخل افزایش می‌یابد [۱۲، ۱۳]. ویژگی‌های ترشوندگی سنگ نیز بر توزیع سیال تأثیر می‌گذارد؛ به‌طور معمول، فاز ترشونده در منافذ ریزتر قرار می‌گیرد و جریان فاز غیرترشونده را محدود می‌سازد [۵]. کشش بین‌سطحی میان فازهای سیال بر نیروهای موینگی و کارایی جابه‌جایی تأثیر دارد؛ به‌ویژه در شرایط نزدیک به نقطه بحرانی که کشش بین‌سطحی کاهش می‌یابد، رفتار نفوذپذیری نسبی ممکن است از حالت امتزاج ناپذیر به حالت امتزاج پذیر تغییر یابد [۱۴].

علاوه بر این، هندسه و توپولوژی شبکه منفذی شامل توزیع اندازه منافذ، پیوستگی^۲ و پیچیدگی^۳ نقش مهمی در نفوذپذیری نسبی ایفا می‌کنند. روش‌هایی مانند تحلیل تصویری مقاطع نازک^۴ برای توصیف ساختار منفذی^۵ و پیش‌بینی رفتار جریان مفید هستند [۱۵، ۱۶]. پدیده هیستریزیس نیز در نفوذپذیری نسبی مشاهده می‌شود، به‌طوری‌که رابطه آن با اشباع در فرآیندهای تغذیه^۶ و آشام^۷ متفاوت است، که عمدتاً ناشی از اثرات بین‌سطحی و به دام افتادن فازهای سیال در فضای منفذی است [۱۷].

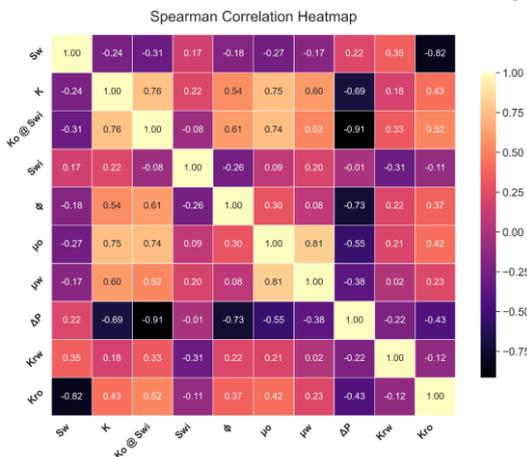
نفوذپذیری نسبی معمولاً به‌صورت آزمایشگاهی با استفاده از مغزه‌های استخراج‌شده از مخازن یا سایر مواد متخلخل اندازه‌گیری می‌شود [۵]. روش‌های مختلفی از جمله حالت پایدار^۸ و حالت ناپایدار^۹ برای تعیین منحنی‌های نفوذپذیری نسبی به‌کار می‌روند [۶]. با این حال، این آزمایش‌ها پرهزینه و زمان‌بر هستند [۱۵]. از این‌رو، مدل‌های متعددی برای پیش‌بینی نفوذپذیری نسبی بر اساس توزیع اندازه منافذ،

Steady-State^۸
Unsteady-State^۹
Percolation Theory^{۱۰}
Machine Learning^{۱۱}
Real-Time^{۱۲}
Upscaling^{۱۳}
Dual-Wettability^{۱۴}

Hysteresis^۱
Connectivity^۲
Tortuosity^۳
image analysis of thin sections^۴
pore structures^۵
Imbibition^۶
Drainage^۷

نسبی انتخاب شدند.

استفاده از متغیرهای ورودی متنوع و اندازه‌گیری شده به صورت تجربی از نمونه‌های مغزه واقعی، بنیانی قوی و قابل اعتماد برای توسعه مدل‌های مبتنی بر داده فراهم می‌کند. این رویکرد موجب می‌شود مدل‌های آموزش دیده بر پایه واقعیت‌های فیزیکی بنا شده و قادر به شناسایی روابط پیچیده میان ویژگی‌های سنگ و رفتار جریان سیال در محیط متخلخل باشند. افزون بر این، بهره‌گیری از داده‌های آزمایشگاهی دقت و قابلیت تعمیم‌پذیری مدل‌ها را افزایش داده و نتایج حاصل را برای نواحی مختلف مخزن کاربردی‌تر می‌سازد.



شکل ۱. مقادیر ضریب همبستگی اسپیرمن متغیرها

برای بررسی همبستگی بین پارامترهای ورودی و خروجی، از روش همبستگی اسپیرمن استفاده شد که نتایج آن در شکل ۱ نشان داده شده است. در این روش، مقادیر $SCC = \pm 1$ بیانگر همبستگی قوی بین دو پارامتر و مقادیر صفر نشان‌دهنده عدم وجود رابطه معنادار هستند. همچنین، مقادیر مثبت نشان‌دهنده همبستگی مستقیم و مقادیر منفی نشان‌دهنده همبستگی معکوس بین دو پارامتر می‌باشند [۲۳]. بررسی مقادیر موجود در شکل نشان‌دهنده عدم وجود رابطه قوی بین پارامترهای ورودی و خروجی است که ضرورت استفاده از روش‌های یادگیری ماشین برای مدل‌سازی نفوذپذیری نسبی آب و نفت را نشان می‌دهد. شایان ذکر است که نتایج این بخش آنالیز حساسیت محسوب نمی‌شوند، زیرا مقادیر تا حد زیادی وابسته به پایگاه داده مورد استفاده هستند.

هستند که پیش‌بینی دقیق نفوذپذیری را دشوار می‌سازد. یادگیری ماشین می‌تواند مدل‌های قابل اعتماد برای این محیط‌های پیچیده ایجاد کند [۲۲].

- مخازن شیل: مشخصه‌یابی مخازن شیل به دلیل نفوذپذیری بسیار پایین دشوار است، اما یادگیری ماشین امکان پیش‌بینی سریع، دقیق و مقرون به صرفه نفوذپذیری را در این نوع مخازن فراهم می‌کند [۱۹].

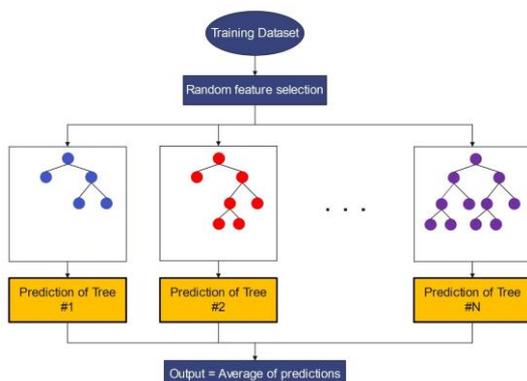
با توجه به بررسی‌های انجام شده، علیرغم وجود مطالعاتی در زمینه برآورد نفوذپذیری نسبی، اغلب تحقیقات یا از روش‌های کلاسیک آزمایشگاهی استفاده کرده‌اند و یا حجم داده‌های ورودی محدود بوده و تحلیل دقیق‌تری از نقش پارامترها در مدل ارائه نداده‌اند. همچنین، مقایسه مستقیم چندین الگوریتم یادگیری ماشین مدرن و تفسیرپذیری مدل از طریق روش SHAP به ندرت در مطالعات گذشته مشاهده شده است. ویژگی متمایز این پژوهش، ترکیب مجموعه‌ای از مدل‌های قدرتمند یادگیری ماشین شامل *Extra Trees*، *KNN*، *CatBoost* و *XGBoost* با داده‌های آزمایشگاهی واقعی و تحلیل تفسیرپذیری مدل است که امکان ارزیابی دقیق عملکرد الگوریتم‌ها در شرایط مختلف مخزن را فراهم می‌سازد. این رویکرد نه تنها دقت مدل‌سازی را افزایش داده، بلکه قابلیت تعمیم‌پذیری و کاربرد صنعتی نتایج را نیز ارتقا می‌دهد.

۲. روش‌شناسی

۲.۱. جمع‌آوری و پیش‌پردازش داده

در این پژوهش، داده‌ها از ۱۶ نمونه مغزه استخراج شده از یک مخزن نفتی واقع در جنوب غرب ایران گردآوری شدند. این نمونه‌ها از عمق ۴۲۴۴.۲۴ تا ۴۲۴۷.۹۱ متری سازند فلهلیان (لیتولوژی: کربناته) جمع‌آوری شده‌اند. تا نوع زمین‌شناسی و پتروفیزیکی سنگ به طور مناسبی نمایان شود. بر روی مغزه‌ها آزمایش‌های آزمایشگاهی به منظور اندازه‌گیری خواص اساسی سنگ و سیال در دمای محیط انجام گرفت. این مجموعه داده شامل ۴۲۸ نقطه داده بوده و متغیرهای مورد استفاده برای مدل‌سازی عبارت‌اند از: نفوذپذیری مطلق، نفوذپذیری نفت در اشباع آب غیرقابل کاهش (*Swi*)، اشباع آب غیرقابل کاهش، تخلخل، گرانیوی نفت، گرانیوی آب و اختلاف فشار. این پارامترها به عنوان ویژگی‌های ورودی برای آموزش مدل‌های یادگیری ماشین در پیش‌بینی نفوذپذیری

تأثیر قرار دهد. این رویکرد برای مسائل دارای تعداد زیادی ویژگی پیوسته مناسب است، زیرا اغلب دقت را افزایش داده و هزینه محاسباتی یافتن تقسیم‌های بهینه را کاهش می‌دهد که در درخت‌های تصمیم استاندارد معمولاً لازم است. در حالت کاملاً تصادفی، *Extra Trees* درختانی می‌سازد که ساختار آن‌ها کاملاً مستقل از مقادیر خروجی داده‌های آموزش است. سطح تصادفی‌سازی می‌تواند از طریق یک پارامتر تنظیم شود تا با انواع مختلف مسائل سازگار باشد.



شکل ۲: شماتیک مدل Extra trees [۲۴]

این الگوریتم از رویکرد معمول ساخت درخت از بالا به پایین^۶ استفاده می‌کند، اما در دو جنبه با روش‌های سنتی درختی تفاوت دارد: نقاط تقسیم به صورت تصادفی انتخاب می‌شوند و کل مجموعه داده‌های آموزش برای رشد هر درخت به کار گرفته می‌شود. پیش‌بینی نهایی از طریق ترکیب خروجی تمام درختان حاصل می‌شود—با استفاده از رأی اکثریت^۸ () در مسائل دسته‌بندی و میانگین‌گیری در مسائل رگرسیون. از منظر تعادل خطا-واریانس^۹، *Extra Trees* با تصادفی‌سازی گسترده و میانگین‌گیری تجمعی واریانس را کاهش می‌دهد که می‌تواند عملکرد تعمیم‌پذیری^{۱۰} مدل را بهبود بخشد. از نظر محاسباتی، با فرض تعادل در درخت‌ها، پیچیدگی آموزش تقریباً برابر $O(N \log N)$ است که N اندازه مجموعه داده می‌باشد. رفتار مدل عمدتاً توسط پنج پارامتر اصلی هدایت

برای دستیابی به درک بهتر از داده‌ها، پارامترهای آماری هر متغیر در ارائه شده است. جدول ۱ خلاصه آماری نشان‌دهنده یک سامانه مخزنی ناهمگن است که با تغییرپذیری قابل توجه و چولگی مثبت^۱ در اغلب خواص، به‌ویژه نفوذپذیری و نفوذپذیری نسبی، مشخص می‌شود. علاوه بر این، دامنه وسیع تمامی پارامترها می‌تواند قابلیت تعمیم‌پذیری مدل را افزایش دهد.

جدول ۱. پارامترهای آماری متغیرها

مشخصات	میانگین	بیشترین مقدار	کمترین مقدار	انحراف معیار	چولگی
اشباع آب	۵۰.۳۰	۸۶.۹۹	۱۶.۰۲	۱۵.۷۴	۰.۱۶
نفوذپذیری مطلق	۲۷.۲۵	۹۶.۷۰	۱.۰۷	۲۶.۹۳	۱.۲۹
نفوذپذیری نفت در S_{wi}	۴.۵۶	۲۴.۲۷	۰.۰۰	۶.۳۹	۱.۶۹
اشباع آب غیرقابل کاهش	۲۸.۶۶	۴۸.۳۱	۱۶.۰۳	۹.۹۱	۰.۶۴
تخلخل	۱۹.۳۵	۲۸.۱۳	۱۱.۵۶	۴.۸۷	۰.۲۳
گرانروی نفت	۹.۰۱	۱۹	۵.۴۷	۵.۹۵	۱.۰۸
گرانروی آب	۱.۲۸	۱.۵۵	۱.۱۷	۰.۱۶	۱.۰۶
اختلاف فشار	۳۹۵.۶۷	۱۵۰۰.۲	۲۰	۴۸۲.۷۳	۱.۲۴
نفوذپذیری نسبی آب	۰.۰۷	۰.۶۱	۰	۰.۰۹	۲.۴۷
نفوذپذیری نسبی نفت	۰.۱۵	۰.۹۹	۰.۰۰	۰.۲۲	۱.۸۰

۲.۲. یادگیری ماشین

۱.۲.۲ Extra trees

الگوریتم *Extra Trees* یک روش یادگیری تجمعی مبتنی بر الگوریتم‌های درخت تصمیم^۲ است که برای مسائل دسته‌بندی^۳ و رگرسیون^۴ مناسب می‌باشد. همان‌گونه که در شکل ۲ مشاهده می‌شود، ایده اصلی استفاده از درختان تصادفی- به‌ویژه در حضور ویژگی‌های عددی پیوسته- این است که انتخاب نقاط تقسیم^۵ به صورت تصادفی می‌تواند تغییرپذیری^۶ در درخت حاصل را به‌طور قابل‌توجهی تحت

^۶ Variability
^۷ Top-Down Tree-Building
^۸ Majority Voting
^۹ Bias-Variance
^{۱۰} Generalization

^۱ Positive Skewness
^۲ Decision Tree
^۳ Classification
^۴ Regression
^۵ Split Points

می‌شود:

- $max_features$: تعداد ویژگی‌های انتخاب‌شده به صورت تصادفی در هر گره، که تنوع تقسیم‌ها را تعیین می‌کند،
 - $min_samples_leaf$: حداقل تعداد نمونه لازم برای تشکیل یک گره برگ، که هموارسازی مدل را کنترل می‌کند،
 - $min_samples_split$: حداقل تعداد نمونه مورد نیاز برای تقسیم یک گره داخلی، که عمق و پیچیدگی درخت را تحت تأثیر قرار می‌دهد،
 - max_depth : حداکثر عمق هر درخت فردی، که از بیش‌برازش (*Overfitting*) جلوگیری می‌کند،
 - $n_estimators$: تعداد درخت‌ها در مجموعه، که بر پایداری و عملکرد کلی مدل تأثیر دارد.
- ساختار درخت شامل گره ریشه (*Root*)، گره‌های داخلی (*Internal*) و برگ‌ها (*Leaf Nodes*) است که هر یک نقش مهمی در تعیین رفتار پیش‌بینی مدل ایفا می‌کنند [۲۵-۲۸].

می‌کند. این فاصله معمولاً با استفاده از معیارهای فاصله‌ای مانند فاصله منهتن^۴، مینکوفسکی^۵، یا فاصله اقلیدسی^۶ که به‌طور رایج ترجیح داده می‌شود، محاسبه می‌گردد. فاصله اقلیدسی، که در معادله (۱) بیان شده است، برای اندازه‌گیری فاصله مستقیم بین دو نقطه X_i و Y_i در یک فضای n -بعدی به‌کار می‌رود.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (Y_i - X_i)^2} \quad (1)$$

پس از محاسبه تمام مقادیر فاصله، الگوریتم نزدیک‌ترین نقاط داده را بر اساس کمترین مقادیر انتخاب می‌کند. برای برآورد خروجی یک نمونه ناشناخته، معمولاً میانگین مقادیر هدف^۷ متناظر با این نقاط انتخاب‌شده گرفته می‌شود. معادله (۲) برای تعیین این خروجی تخمینی به‌کار می‌رود.

$$\hat{Y} = \frac{1}{k} \sum_{i=1}^k y_i \quad (2)$$

که در آن y_i نمایانگر مقدار هدف همسایه نزدیک‌ترین نقطه i ام است [۲۹-۳۲].

۳.۲.۲. Categorical boosting

الگوریتم *CatBoost* نوعی از الگوریتم‌های گرادیان بوستینگ^۸ است که به‌طور خاص برای مدیریت مؤثر ویژگی‌های دسته‌ای^۹ طراحی شده است. این الگوریتم از درختان تصمیم باینری^{۱۰} به‌عنوان یادگیرنده‌های پایه^{۱۱} استفاده می‌کند، همان‌طور که در شکل ۳ نشان داده شده است. یکی از نقاط قوت اصلی *CatBoost* در کاهش از دست رفتن اطلاعات هنگام کار با متغیرهای دسته‌ای نهفته است.

مفهوم اساسی برای درک عملکرد *CatBoost*، برخورد متمایز آن با داده‌های آموزش و آزمون است که به کاهش نشت هدف^{۱۲} و بیش‌برازش کمک می‌کند.

۲.۲.۲. K-Nearest Neighbor

الگوریتم *K-Nearest Neighbors (KNN)* یک روش رگرسیون است که بدون فرض هیچ توزیع خاصی برای داده‌ها عمل می‌کند و در دسته یادگیری مبتنی بر نمونه^۱ قرار می‌گیرد. این الگوریتم پیش‌بینی‌ها را با شناسایی K مشاهده مشابه‌ترین نقاط در مجموعه آموزش بر اساس یک معیار فاصله مشخص انجام می‌دهد. پارامتر K نقش مهمی در دقت مدل دارد؛ انتخاب مقدار نامناسب می‌تواند منجر به بیش‌برازش^۲ یا کم‌برازش^۳ مدل شود. این روش شباهت را با اندازه‌گیری فاصله بین نمونه ورودی و سایر نقاط مجموعه داده تعیین می‌کند و امکان استنباط خروجی از نزدیک‌ترین همسایگان را فراهم می‌سازد.

فرض کنید مجموعه داده‌ای شامل ویژگی‌های ورودی X و برچسب‌های خروجی متناظر Y با تعداد n نقطه داده موجود باشد. برای پیش‌بینی یک نمونه جدید x الگوریتم *KNN* فاصله بین x و هر نقطه موجود در مجموعه X را محاسبه

Target Values^۷
Gradient Boosting^۸
Categorical Features^۹
Binary Decision Trees^{۱۰}
Base Learners^{۱۱}
Target Leakage^{۱۲}

Instance-Based Learning^۱
Overfitting^۲
Underfitting^۳
Manhattan^۴
Minkowski^۵
Euclidean^۶

پیش‌بینی شده در این فرآیند با استفاده از رابطه زیر محاسبه می‌شود:

$$T = H(x_i) = \sum_{k=1}^n c_n l_{\{x \in R_n\}} \quad (4)$$

که در آن H نشان‌دهنده تابع درخت تصمیم، x_i متغیر توضیحی^۳ و R_n بخش نامتصل مرتبط با برگ‌های درخت^۴ است [۳۳-۳۶].

۴.۲.۲. Extreme gradient boosting

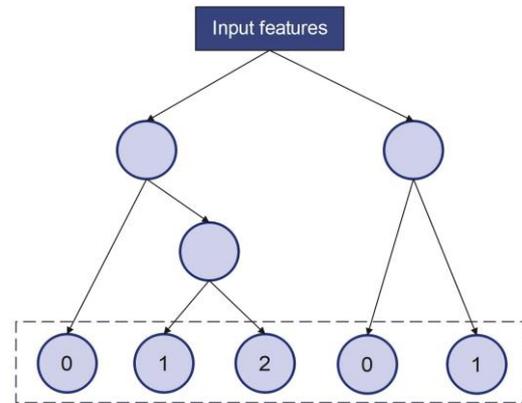
الگوریتم $XGBoost$ یک تکنیک یادگیری نظارت‌شده^۵ مبتنی بر گرادین بوستینگ^۶ است که با ترکیب تعداد زیادی درخت تصمیم و رگرسیون، مدلی پیش‌بینی‌کننده قوی و تعمیم‌پذیر ایجاد می‌کند (*Error! Reference source not found.*)ی ک مدل $CART$ شامل یک گره شروع، چندین نقطه تصمیم‌گیری و گره‌های خروجی نهایی است. برای مجموعه داده‌ای شامل n نمونه و m ویژگی ورودی، $XGBoost$ پیش‌بینی نهایی خود را با تجمیع خروجی‌های K مدل $CART$ مجزا انجام می‌دهد که به صورت ریاضی به شکل زیر بیان می‌شود:

$$\hat{y}_i = \sum_{k=1}^K f_k(X_i), f_k \in F \quad (5)$$

در این فرمول‌بندی، f_k نشان‌دهنده درخت تصمیم k ام و \hat{y}_i نمایانگر نتیجه پیش‌بینی شده توسط مدل $XGBoost$ است. فضای توابع F به دامنه‌ای اشاره دارد که تمام ساختارهای ممکن $CART$ را شامل می‌شود.

$$F = \{f(X) = \omega_q(x)\}, \quad (q: R^m \rightarrow T, \omega \in R^T) \quad (6)$$

در اینجا، q تابع تصمیمی است که هر ورودی X را به یک شاخص برگ^۷ مشخص در درخت نگاشت می‌کند؛ $\omega_q(x)$ مقدار برگ اختصاص‌یافته به آن شاخص را نشان می‌دهد؛ T تعداد کل برگ‌های درخت است؛ و $\omega \in R^T$ بردار وزن‌های برگ‌ها را تشکیل می‌دهد.



مفهوم کلیدی دیگر در درک نحوه کدگذاری ویژگی‌های دسته‌ای توسط $CatBoost$ ، استفاده از تابع شاخص l است که نقش محوری در فرآیند کدگذاری ایفا می‌کند، همان‌طور که در ادامه نشان داده شده است.

$$Indicator Function l_{k,t} = \begin{cases} 1 & \text{if } k = t \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

تابع شاخص که در بالا توضیح داده شد، نقش کلیدی در فرمولی ایفا می‌کند که $CatBoost$ برای تبدیل داده‌های دسته‌ای به مقادیر عددی استفاده می‌کند. فرآیند بوستینگ $CatBoost$ به‌طور ویژه برای مدیریت مؤثر ویژگی‌های دسته‌ای از طریق تکنیک‌های پردازش تخصصی طراحی شده است. از مهم‌ترین این تکنیک‌ها می‌توان به آمار مبتنی بر هدف^۱ و کدگذاری $One-Hot-Max-Size (OHMS)$ اشاره کرد. مراحل اصلی روش $CatBoost$ شامل ایجاد زیرمجموعه تصادفی از داده‌ها، تبدیل برچسب‌های دسته‌ای به نمایش‌های عددی و تبدیل ویژگی‌ها به قالب عددی مناسب برای مدل‌سازی است.

برای کاهش خطر بیش‌برازش، $CatBoost$ هنگام برآورد مقادیر برگ‌ها^۲ در طول ساخت درخت، جایگشت‌های تصادفی از داده‌های آموزش اعمال می‌کند. این مکانیزم منحصر به فرد موجب افزایش پایداری و مقاومت مدل می‌شود. مقدار

^۵ Supervised Learning

^۶ Gradient Boosting

^۷ Leaf Index

^۱ Target-Based Statistics

^۲ Leaf Values

^۳ Explanatory Variable

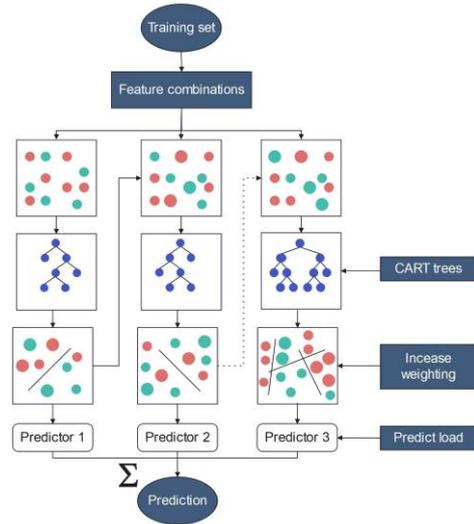
^۴ Tree's Leaves

برای توسعه پیش‌بینی $f(x)$ تابع هدف باید با کمینه‌سازی بهینه‌سازی شود.

$$\mathcal{L}(\phi) = \sum_l^n l(\hat{y}_l, y_l) + \sum_k^K \Omega(f_k) \quad (7)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (8)$$

در این فرمول‌بندی، l به یک تابع خطای محدب و قابل مشتق اشاره دارد، در حالی که Ω مؤلفه منظم‌سازی^۱ است که پیچیدگی مدل را جریمه می‌کند. پارامتر γ سهم تابع خطا را تنظیم می‌کند، λ قدرت منظم‌سازی را کنترل می‌نماید و ω وزن اختصاص‌یافته به هر برگ درخت را نشان می‌دهد.



شکل ۴: شماتیک مدل XGBoost [۳۷]

جزء اول در معادله (۷) خطای پیش‌بینی را بیان می‌کند که معمولاً در مسائل رگرسیون با میانگین مربعات خطا^۲ اندازه‌گیری می‌شود و به مدل اجازه می‌دهد خطا را در طول آموزش کاهش دهد. جزء دوم بر اساس ساختار و عمق هر مدل $CART$ جریمه‌ای اعمال می‌کند و از پیچیدگی بیش از حد جلوگیری می‌نماید. برای کاهش بیش‌برازش، $XGBoost$ از استراتژی‌های مختلفی مانند تنظیم ابرپارامترها^۳، افزودن مؤلفه‌های منظم‌سازی و اعمال معیارهای توقف زودهنگام^۴ در

طول آموزش استفاده می‌کند.

پس از پیش‌پردازش داده‌ها، ۷۰٪ از داده‌ها به صورت تصادفی برای آموزش مدل‌ها انتخاب شدند. در طول فرآیند آموزش، الگوریتم بیزین^۵ برای بهینه‌سازی ابرپارامترها به کار گرفته شد. برای جلوگیری از بیش‌برازش روی داده‌های آموزش، از تکنیک اعتبارسنجی متقابل ۵-تایی^۶ استفاده شد. پس از ساخت مدل‌ها، دقت آن‌ها بر روی داده‌های دیده‌نشده با استفاده از ۳۰٪ باقی‌مانده از داده‌ها ارزیابی گردید.

۳. نتایج و توضیحات

۱.۱.۳. ارزیابی مدل

۱.۱.۳.۱. نفوذپذیری نسبی آب

مقادیر بهینه ابرپارامترها در ساخت مدل‌های نفوذپذیری نسبی آب با استفاده از روش بیزین در جدول ۲ ارائه شده است.

جدول ۲. مقادیر بهینه ابرپارامترهای مدل‌های مربوط به نفوذپذیری نسبی آب

Methods	Hyperparameter	Value
CatBoost	depth	۸.۰۰۰۰
	l2_leaf_reg	۲.۸۹۶۷
	learning rate	۰.۰۸۹۱
XGBoost	max_depth	۸.۰۰۰۰
	learning rate	۰.۳۴۳۵
	n_estimators	۳۰۷.۰۰۰۰
	gamma	۰.۰۰۰۰
	min_child_weight	۸.۰۱۶۱
	subsample	۰.۶۴۸۴
Extra trees	colsample_bytree	۱.۰۰۰۰
	max_depth	۲۰.۰۰۰۰
	max_features	۰.۴۶۲۶
	min_samples_leaf	۱.۰۰۰۰
	min_sample_split	۲.۰۰۰۰
KNN	n_estimators	۳۹۷.۰۰۰۰
	n_neighbors	۲.۰۰۰۰
	p	۱.۰۰۰۰

^۴ Early Stopping

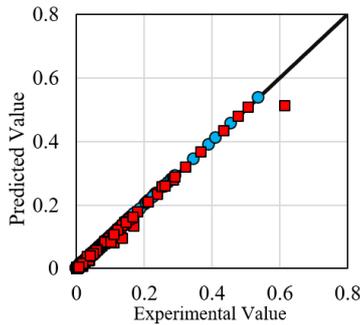
^۵ Bayesian

^۶ Fold Cross Validation

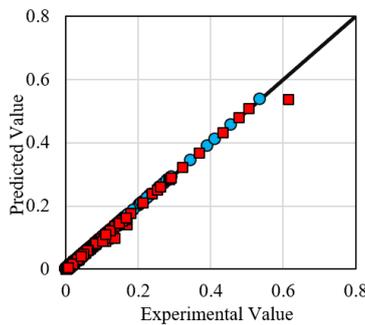
^۱ Regularization

^۲ Mean Squared Error

^۳ Hyperparameter Tuning



شکل ۷: نمودار تقاطعی مدل KNN مربوط به نفوذپذیری نسبی آب

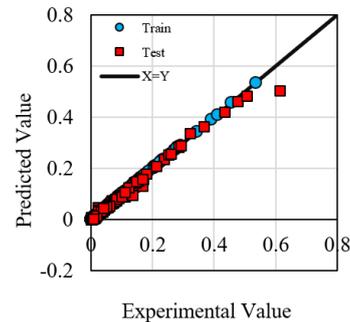


شکل ۸: نمودار تقاطعی مدل Extra tree مربوط به نفوذپذیری نسبی آب

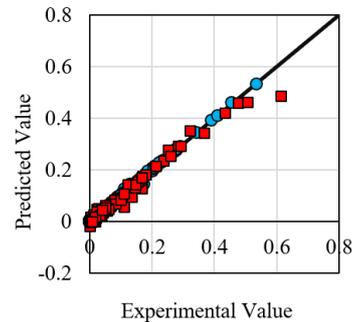
برای ارزیابی دقیق‌تر عملکرد مدل‌ها، از معیارهای خطا استفاده شد. این معیارها شامل ضریب تعیین^۲، ریشه میانگین مربعات خطا (*Root Mean Square Error, RMSE*) و میانگین قدر مطلق خطا (*Mean Absolute Error, MAE*) هستند. همان‌طور که در *Error! Reference source not found* نشان داده شده است، با بررسی مقادیر R^2 می‌توان نتیجه گرفت که تمامی مدل‌ها در مجموعه داده آموزش عملکرد بسیار خوبی داشته‌اند. با این حال، همان‌طور که از نمودارهای متقاطع انتظار می‌رفت، روش *XGBoost* با $R^2 = 0.9901$ در مقایسه با سایر مدل‌ها عملکرد چندان مناسبی نداشت. دو روش *Extra Trees* و *KNN* در مجموعه آموزش به ضریب تعیین ۱ دست یافتند که بالاترین مقدار در بین مدل‌ها است.

با بررسی داده‌های مربوط به ضریب تعیین مجموعه آزمون، روش *Extra Trees* با $R^2 = 0.9941$ بیشترین دقت را نشان داد و پس از آن روش *KNN* با $R^2 = 0.9902$ قرار گرفت.

شکل ۵ تا شکل ۸ نمودار متقاطع^۱ مدل‌های توسعه داده شده برای نفوذپذیری نسبی آب را نشان می‌دهند. در این روش، خروجی‌های مدل بر اساس مقادیر واقعی ترسیم می‌شوند و هر چه نقاط به خط ایده‌آل نزدیک‌تر باشند، عملکرد مدل بهتر خواهد بود. مطابق این شکل، تمامی روش‌ها عملکرد خوبی در برآورد نفوذپذیری نسبی آب نشان دادند. با این حال، به نظر می‌رسد روش *XGBoost* (شکل ۶) در مقایسه با سایر روش‌ها عملکرد مناسبی نداشته و پراکندگی نقاط در نمودار مربوط به این روش نسبت به دیگر روش‌ها بیشتر است. همچنین، به نظر می‌رسد روش *CatBoost* (شکل ۵) در برآورد مقادیر پایین نفوذپذیری نسبی آب موفقیت چندان نداشته است. درباره دو روش دیگر به ترتیب شکل ۷ و شکل ۸، یعنی *Extra Trees* و *KNN*، نمی‌توان اظهار نظر قطعی داشت، اما به نظر می‌رسد روش *Extra Trees* در بازه داده‌ای ۰.۱ تا ۰.۲ موفق‌تر بوده و در سایر بازه‌ها رفتار دو مدل بسیار مشابه یکدیگر است. برای تحلیل دقیق‌تر، استفاده از معیارهای خطا ضروری است.



شکل ۵: نمودار تقاطعی مدل CatBoost مربوط به نفوذپذیری نسبی آب



شکل ۶: نمودار تقاطعی مدل XGBoost مربوط به نفوذپذیری نسبی آب

^۲ Coefficient of Determination

^۱ Cross-Plot

XGBoost قرار دارند.

۲.۱.۳. نفوذپذیری نسبی نفت

مقادیر بهینه ابرپارامترها در ساخت مدل‌های نفوذپذیری نسبی نفت با استفاده از روش بیزین در جدول ۴ ارائه شده است.

جدول ۴. مقادیر بهینه ابرپارامترهای مدل‌های مربوط به

نفوذپذیری نسبی نفت		
Methods	Hyperparameter	Value
CatBoost	depth	۶.۰۰۰۰
	l2_leaf_reg	۱.۴۶۸۸
	learning_rate	۰.۰۸۶۵
	max_depth	۳.۰۰۰۰
XGBoost	learning_rate	۰.۵۰۰۰
	n_estimators	۴۶۰.۰۰۰۰
	gamma	۰.۰۰۰۰
	min_child_weight	۱۰.۰۰۰۰
	subsample	۱.۰۰۰۰
Extra trees	colsample_bytree	۱.۰۰۰۰
	max_depth	۱۶.۰۰۰۰
	max_features	۱.۰۰۰۰
	min_samples_leaf	۹.۰۰۰۰
KNN	min_sample_split	۲۵۰.۰۰۰۰
	n_estimators	۳۹۷.۰۰۰۰
	n_neighbors	۲.۰۰۰۰
	p	۱.۰۰۰۰

بررسی نمودارهای متقاطع مدل‌های نفوذپذیری نسبی نفت (شکل ۹ تا شکل ۱۲) نشان می‌دهد که برآورد این پارامتر برای مدل‌ها نسبت به نفوذپذیری نسبی آب دشوارتر است. این پدیده احتمالاً به دلیل پیچیدگی بالای روابط بین متغیرهای ورودی و خروجی رخ می‌دهد. برخلاف مدل‌های نفوذپذیری نسبی آب، روش *Extra Trees* (شکل ۱۲) ضعیف‌ترین عملکرد را در بین مدل‌ها دارد و تقریباً در هیچ بازه‌ای از داده‌های نفوذپذیری نسبی نفت عملکرد مناسبی ارائه نکرده است. بررسی نمودار روش *XGBoost* (شکل ۱۰) نشان می‌دهد که این مدل نیز مانند قبل در برآورد خروجی موفق نبوده است، با این حال، نقاط نمودار این مدل، به‌ویژه در مقادیر نفوذپذیری بالاتر از ۰.۴، بیشتر به خط ایده‌آل نزدیک هستند. عملکرد دو روش *KNN* و *CatBoost* (به ترتیب شکل ۹ و شکل ۱۱) به یکدیگر نزدیک بوده است، اما

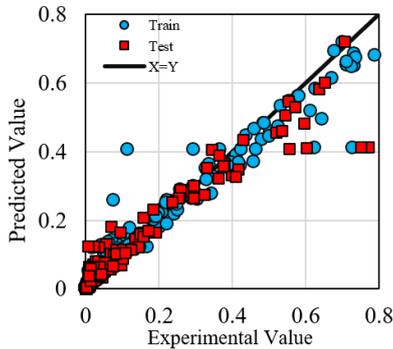
این تفاوت نشان می‌دهد که اگرچه هر دو روش توانسته‌اند مجموعه آموزش را به‌طور کامل یاد بگیرند، روش *Extra Trees* قابلیت تعمیم‌پذیری بالاتری نسبت به داده‌های جدید دارد. همچنین، در مجموعه آزمون، مشابه مجموعه آموزش، ضعیف‌ترین عملکرد به روش *XGBoost* تعلق دارد. روش *CatBoost* نیز در مجموعه آزمون ضریب تعیین ۰.۹۸۸۰ را نشان داد که کاهش قابل توجهی نسبت به مجموعه آموزش دارد. روند ضریب تعیین در کل داده‌ها نیز مشابه است؛ روش *Extra Trees* با $R^2 = ۰.۹۹۷۴$ بهترین مدل است و پس از آن *KNN*، *CatBoost* و *XGBoost* با مقادیر ضریب تعیین ۰.۹۹۵۸، ۰.۹۹۴۸ و ۰.۹۸۹۸ قرار دارند.

جدول ۳. معیارهای خطای مدل‌های ارائه شده مرتبط با

نفوذپذیری نسبی آب					
Metric	Subset	CatBoost	XGBoost	KNN	Extra trees
R2	Train	۰.۹۹۹	۰.۹۹۰	۱.۰۰۰	۱.۰۰۰
	Test	۰.۹۸۸	۰.۹۷۸	۰.۹۹۰	۰.۹۹۴
	Overall	۰.۹۹۵	۰.۹۹۰	۰.۹۹۸	۰.۹۹۷
RMSE	Train	۰.۰۰۱	۰.۰۰۳	۰.۰۰۰	۰.۰۰۰
	Test	۰.۰۱۲	۰.۰۱۶	۰.۰۱۱	۰.۰۰۸
	Overall	۰.۰۰۷	۰.۰۰۹	۰.۰۰۶	۰.۰۰۵
MAE	Train	۰.۰۰۱	۰.۰۰۱	۰.۰۰۰	۰.۰۰۰
	Test	۰.۰۰۵	۰.۰۰۹	۰.۰۰۳	۰.۰۰۲
	Overall	۰.۰۰۲	۰.۰۰۳	۰.۰۰۱	۰.۰۰۱

بررسی مقادیر *RMSE* و *MAE* نیز با داده‌های ضریب تعیین هم‌راستا است. در مجموعه آموزش، روش‌های *Extra Trees* و *KNN* مقادیر *RMSE* و *MAE* برابر صفر دارند که نشان‌دهنده تسلط کامل این دو مدل بر داده‌های آموزش است. در مجموعه آزمون، روش *Extra Trees* دارای *RMSE* و *MAE* به ترتیب برابر ۰.۰۰۸۳ و ۰.۰۰۲۲ است که نشان می‌دهد این مدل در مقایسه با روش *KNN* با *RMSE* و *MAE* برابر ۰.۰۱۰۶ و ۰.۰۰۳۲ قابلیت تعمیم بالاتری دارد و در مواجهه با داده‌های جدید عملکرد بهتری ارائه می‌کند. همچنین، روش *XGBoost* با *RMSE* و *MAE* کلی ۰.۰۰۹۱ و ۰.۰۰۳۳ ضعیف‌ترین عملکرد را نشان می‌دهد.

به‌طور کلی، با بررسی نمودارهای متقاطع و معیارهای خطا می‌توان نتیجه گرفت که روش *Extra Trees* به دلیل قابلیت تعمیم بالاتر، بهترین روش برای برآورد نفوذپذیری نسبی آب است و پس از آن به ترتیب روش‌های *KNN*، *CatBoost* و



شکل ۱۲: نمودار تقاطعی مدل Extra trees مربوط به نفوذپذیری نسبی نفت

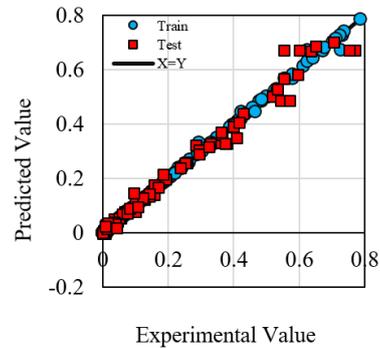
نتایج معیارهای خطا مربوط به مدل‌های برآورد نفوذپذیری نسبی نفت (جدول ۵) نشان می‌دهد که به طور کلی، عملکرد مدل‌ها در این بخش به خوبی مدل‌های نفوذپذیری نسبی آب نبوده است. روش *KNN*، همانند حالت قبلی، مقدار ضریب تعیین (R^2) برابر با ۱ را بر روی داده‌های آموزشی نشان می‌دهد که بیانگر یادگیری کامل مدل است. همچنین، این مدل بر روی داده‌های آزمایشی مقدار ضریب تعیین برابر با ۰.۹۹۰۷ را داشته است که نشان‌دهنده توانایی بسیار خوب مدل در تعمیم‌دهی به داده‌های جدید است. مقدار خطای میانگین مربعات (*RMSE*) و میانگین قدر مطلق خطا (*MAE*) برای این مدل در داده‌های آموزشی برابر با صفر بوده است. علاوه بر این، در مجموعه داده آزمایشی، مقادیر *RMSE* و *MAE* به ترتیب برابر با ۰.۰۲۰۶ و ۰.۰۰۸۱ گزارش شده‌اند.

جدول ۵. معیارهای خطای مدل‌های ارائه شده مرتبط با

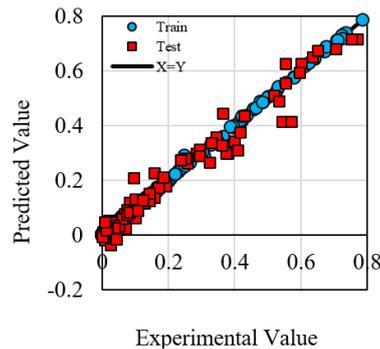
نفوذپذیری نسبی نفت					
Metric	Subset	CatBoost	XGBoost	KNN	Extra trees
R2	Train	۰.۹۹۹	۰.۹۹۹	۱.۰۰۰	۰.۹۷۷
	Test	۰.۹۸۸	۰.۹۷۷	۰.۹۹۱	۰.۹۳۸
	Overall	۰.۹۹۶	۰.۹۹۳	۰.۹۹۷	۰.۹۶۷
RMSE	Train	۰.۰۰۶	۰.۰۰۴	۰.۰۰۰	۰.۰۳۴
	Test	۰.۰۲۴	۰.۰۳۳	۰.۰۲۱	۰.۵۳۳
	Overall	۰.۰۱۴	۰.۰۱۸	۰.۰۱۱	۰.۰۴۱
MAE	Train	۰.۰۰۳	۰.۰۰۳	۰.۰۰۰	۰.۰۱۳
	Test	۰.۰۱۲	۰.۰۱۹	۰.۰۰۸	۰.۰۲۳
	Overall	۰.۰۰۵	۰.۰۰۸	۰.۰۰۲	۰.۰۱۶

دو روش *XGBoost* و *CatBoost* عملکردی بسیار مشابه بر روی داده‌های آموزشی داشته‌اند، اما روش *XGBoost* با مقدار

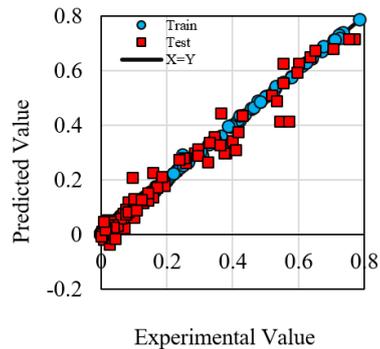
واضح است که روش *KNN* نسبت به *CatBoost* موفق‌تر عمل کرده است. این نکته به‌ویژه در رابطه با داده‌های کمتر از ۰.۴ به‌صورت قابل توجهی مشاهده می‌شود.



شکل ۹: نمودار تقاطعی مدل CatBoost مربوط به نفوذپذیری نسبی نفت



شکل ۱۰: نمودار تقاطعی مدل XGBoost مربوط به نفوذپذیری نسبی نفت

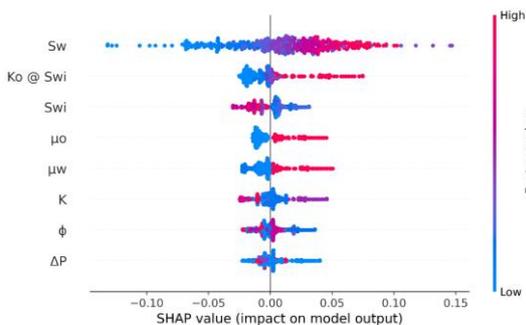


شکل ۱۱: نمودار تقاطعی مدل KNN مربوط به نفوذپذیری نسبی نفت

نسبی آب نشان می‌دهد، اما در مقادیر متوسط و بالا، این رابطه به صورت مستقیم است. پس از اشباع آب، بیشترین تأثیر مربوط به نفوذپذیری نفت در حالت Swi است. این پارامتر نیز در مقادیر پایین و متوسط رابطه‌ای معکوس با نفوذپذیری نسبی آب دارد، در حالی که در مقادیر بالا، این رابطه مستقیم می‌شود. لازم به ذکر است که تأثیر این پارامتر در مقادیر بالا به مراتب بیشتر از مقادیر پایین و متوسط است.

میزان تأثیر شش پارامتر دیگر تقریباً مشابه است، اما نحوه اثرگذاری آن‌ها متفاوت می‌باشد. اشباع اولیه آب در مقادیر پایین رابطه‌ای مستقیم با نفوذپذیری نسبی آب دارد، در حالی که در مقادیر متوسط و بالا این رابطه معکوس می‌شود. همچنین، ویسکوزیته آب و نفت رفتار مشابهی از خود نشان می‌دهند؛ به گونه‌ای که در مقادیر پایین، رابطه‌ای معکوس با نفوذپذیری نسبی دارند و با افزایش مقدار، این رابطه به مستقیم تغییر می‌یابد.

شایان ذکر است که سه پارامتر دیگر رفتار معناداری از خود نشان نمی‌دهند و نمی‌توان به طور قطعی نتیجه‌گیری کرد که هر یک از آن‌ها در چه جهتی با پارامتر خروجی (نفوذپذیری نسبی آب) ارتباط دارند.



شکل ۱۳: نمودار SHAP مدل Extra Trees برای نفوذپذیری نسبی آب

با توجه به اینکه تحلیل حساسیت مبتنی بر روش $SHAP$ برای مدل‌های KNN قابل اعمال نیست، در این پژوهش از مدل $CatBoost$ که از دقت قابل قبولی نیز برخوردار بود (R^2 برابر با ۰.۹۹۶۰)، برای بررسی اثر پارامترهای مختلف بر نفوذپذیری نسبی نفت استفاده شد. نتایج در شکل ۱۴ ارائه شده‌اند. بر اساس این نتایج، اشباع آب بیشترین تأثیر را بر نفوذپذیری نسبی نفت دارد، به گونه‌ای که در مقادیر پایین،

R^2 برابر با ۰.۹۹۹۶ اندکی بهتر از $CatBoost$ با مقدار R^2 برابر با ۰.۹۹۹۲ بوده است. با این حال، در داده‌های آزمایشی این وضعیت معکوس می‌شود، به طوری که روش $CatBoost$ با R^2 برابر با ۰.۹۸۷۹ نسبت به $XGBoost$ با R^2 برابر با ۰.۹۷۶۵ برتری قابل توجهی نشان می‌دهد که بیانگر توانایی بالاتر آن در تعمیم به داده‌های مشاهده‌نشده است. سایر معیارهای خطا نیز با نتایج مربوط به ضریب تعیین سازگار هستند. مقادیر $RMSE$ و MAE برای روش $CatBoost$ به ترتیب برابر با ۰.۰۰۶۰ و ۰.۰۰۲۵ بوده‌اند که تقریباً مشابه روش $XGBoost$ با مقادیر $RMSE$ و MAE برابر با ۰.۰۰۴۲ و ۰.۰۰۲۵ است. با این حال، در مجموعه داده آزمایشی، مقادیر $RMSE$ و MAE برای روش $CatBoost$ به ترتیب به ۰.۰۲۳۵ و ۰.۰۱۱۷ می‌رسند که هر دو نشان‌دهنده کاهش خطا نسبت به روش $XGBoost$ هستند.

روش $Extra Trees$ مقادیر $RMSE$ برابر با ۰.۰۳۳۸ و ۰.۵۳۳۰ را به ترتیب در داده‌های آموزشی و آزمایشی نشان داده است که به وضوح افت عملکرد قابل توجهی را نسبت به سایر روش‌ها نمایش می‌دهد. نتایج مربوط به MAE نیز این مسئله را تأیید می‌کنند.

به طور کلی، با بررسی معیارهای خطا و نمودارهای متقاطع می‌توان نتیجه گرفت که روش KNN نه تنها بهترین عملکرد را در یادگیری نفوذپذیری نسبی نفت داشته، بلکه از نظر قابلیت تعمیم نیز نسبت به سایر روش‌ها برتری داشته است. پس از آن، روش‌های $CatBoost$ ، $XGBoost$ و $Extra Trees$ به ترتیب در رتبه‌های بعدی قرار می‌گیرند. در نهایت، می‌توان بیان کرد که با در نظر گرفتن تمامی عوامل و بر اساس مجموعه داده مورد استفاده در این پژوهش، روش $Extra Trees$ برای برآورد نفوذپذیری نسبی آب و روش KNN برای برآورد نفوذپذیری نسبی نفت مناسب‌ترین عملکرد را در میان روش‌های مورد بررسی داشته‌اند.

۲.۳. آنالیز حساسیت

به منظور بررسی اثر هر پارامتر و انجام تحلیل حساسیت، از روش $SHAP$ استفاده شد که نتایج آن در شکل ۱۳ ارائه گردیده است. با بررسی این شکل مشاهده می‌شود که اشباع آب بیشترین تأثیر را بر نفوذپذیری نسبی آب دارد؛ به طوری که در مقادیر پایین، این پارامتر رابطه‌ای معکوس با نفوذپذیری

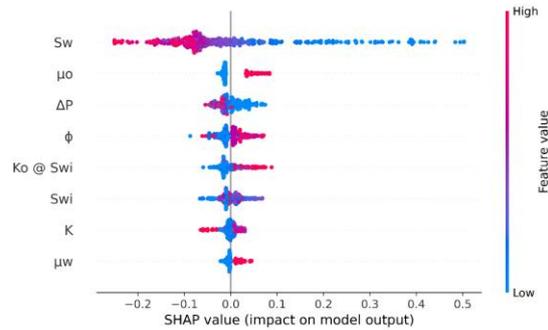
ادغام تکنیک‌های یادگیری ماشین در برآورد منحنی‌های نفوذپذیری نسبی پتانسیلی تحول‌آفرین برای مهندسی مخزن دارد. نفوذپذیری نسبی یکی از ورودی‌های کلیدی در مدل‌های شبیه‌سازی مخزن است که رفتار جریان چندفازی را کنترل کرده و به‌طور مستقیم بر پیش‌بینی برداشت نفت، تولید آب و عملکرد کلی مخزن تأثیر می‌گذارد. به‌طور سنتی، این منحنی‌ها از طریق آزمایش‌های مغزه‌گیری آزمایشگاهی به‌دست می‌آیند که اگرچه از دقت بالایی برخوردارند، اما فرایندی زمان‌بر، پرهزینه و محدود از نظر پوشش مکانی به دلیل محدودیت در دسترسی به مغزه‌ها هستند.

یادگیری ماشین با بهره‌گیری از داده‌های در دسترس میدانی مانند نمودارهای چاه‌نگاری^۱، ویژگی‌های مغزه^۲، نرخ‌های تولید^۳ و پروفایل‌های اشباع^۴، این محدودیت‌ها را برطرف می‌کند. این رویکرد به مهندسان امکان می‌دهد پروفایل‌های پیوسته و میدانی نفوذپذیری نسبی را بدون نیاز به نمونه‌برداری فیزیکی گسترده تولید کنند. برای مهندسان مخزن، این امر به معنای در اختیار داشتن ورودی‌هایی دقیق‌تر و به‌موقع‌تر برای مدل‌های شبیه‌سازی بوده و در نتیجه، قابلیت اطمینان پیش‌بینی‌های تولید و طراحی استراتژی‌های بازیافت را افزایش می‌دهد.

مدل‌های یادگیری ماشین همچنین می‌توانند با ورود داده‌های جدید مجدداً آموزش داده شوند، که این موضوع امکان به‌روزرسانی پویا^۵ مدل‌های مخزن و پشتیبانی از مدیریت تطبیقی^۶ را فراهم می‌سازد. این ویژگی به‌ویژه در مخازن ناهمگن یا دارای داده‌های محدود اهمیت دارد، جایی که روش‌های سنتی قادر به درک کامل تغییرات مکانی نیستند. افزون بر این، یادگیری ماشین می‌تواند در آنالیز حساسیت و کمی‌سازی عدم قطعیت^۷ نیز مورد استفاده قرار گیرد و به مهندسان اجازه دهد سناریوهای مختلف توسعه را با حداقل هزینه محاسباتی بررسی کنند.

در کاربردهای ازدیاد برداشت، که در آن شناخت رفتار سیالات تزریقی تحت شرایط مختلف اشباع و ترشوندگی اهمیت دارد، مدل‌های یادگیری ماشین آموزش‌دیده با داده‌های مرتبط

رابطه‌ای مستقیم با این پارامتر نشان می‌دهد، اما در مقادیر متوسط و بالا، این رابطه معکوس می‌شود. دومین پارامتر مؤثر، ویسکوزیته نفت است که در مقادیر پایین، رابطه‌ای معکوس و ضعیف با خروجی دارد، اما در مقادیر بالاتر، این رابطه مستقیم و قوی‌تر می‌شود. پس از ویسکوزیته نفت، شش پارامتر دیگر تأثیری نسبتاً مشابه اما با رفتارهای متفاوت نشان می‌دهند. اختلاف فشار در مقادیر پایین دارای رابطه‌ای مستقیم و در مقادیر بالا دارای رابطه‌ای معکوس با خروجی است. تخلخل، K_{oi} و ویسکوزیته آب رفتار مشابهی دارند؛ به‌گونه‌ای که در مقادیر پایین، رابطه‌ای معکوس و در مقادیر بالاتر، رابطه‌ای مستقیم با خروجی نشان می‌دهند. دو پارامتر اشباع اولیه آب و نفوذپذیری مطلق نیز رابطه معناداری با نفوذپذیری نسبی نفت نشان نمی‌دهند.



شکل ۱۴: نمودار SHAP مدل Extra Trees برای نفوذپذیری نسبی نفت

نتایج آنالیز حساسیت تا حد زیادی هم‌راستا با مطالعات پیشین می‌باشد. در مطالعه‌ای که توسط زرین و همکاران صورت گرفته نیز، اشباع آب بیشترین تأثیر را بر روی هر دو پارامتر نفوذپذیری آب و نفت داشته‌است [۳۸]. مطالعه‌ای دیگر توسط الالی و همکاران نشان داد اشباع آب موثرترین پارامتر بوده است [۳۹].

۳.۳. پیامدها

Dynamic Update^۵
Adaptive Management^۶
Uncertainty Quantification^۷

Well Logs^۱
Core Properties^۲
Production Rates^۳
Saturation Profiles^۴

برای نفوذپذیری نسبی آب، مدل *Extra Trees* بهترین عملکرد کلی را نشان داد. این مدل بر روی کل مجموعه داده مقدار $R^2 = 0.9974$ را به دست آورد و در داده‌های آزمون نیز تعمیم‌پذیری بسیار خوبی با $R^2 = 0.9941$ نشان داد. همچنین مقادیر $RMSE = 0.0083$ و $MAE = 0.0022$ کمترین میزان خطا را در میان مدل‌ها نشان داد که بیانگر دقت برتر آن است. مدل *KNN* نیز عملکرد قابل توجهی داشت، به‌ویژه در بازه‌های نفوذپذیری پایین، اگرچه از نظر تعمیم‌پذیری اندکی ضعیف‌تر از *Extra Trees* بود.

برای نفوذپذیری نسبی نفت، مدل *KNN* بهترین عملکرد را نشان داد. این مدل با یادگیری کامل داده‌های آموزشی ($R^2 = 1.000$) و حفظ دقت بالا در داده‌های آزمون (0.9907)، $R^2 = 0.9907$ ، $RMSE = 0.0206$ و $MAE = 0.0081$) نتایج بسیار مطلوبی ارائه داد. مدل *CatBoost* نیز عملکرد قابل قبولی داشت، در حالی که مدل *Extra Trees* در مدل‌سازی نفوذپذیری نفت، به‌ویژه در مقادیر بالاتر نفوذپذیری، عملکرد ضعیفی از خود نشان داد.

تحلیل حساسیت مبتنی بر مقادیر *SHAP* نشان داد که اشباع آب بیشترین تأثیر را در برآورد نفوذپذیری نسبی هر دو سیال دارد. در مدل آب، نفوذپذیری نفت در اشباع اولیه آب (*Koi*) و اشباع اولیه آب (*Swi*) نیز اثرگذار بودند، در حالی که در مدل نفت، ویسکوزیته نفت (*Oil Viscosity*) دومین عامل مهم بود، به‌ویژه در مقادیر بالا.

در مجموع، نتایج این پژوهش تأیید می‌کند که یادگیری ماشین رویکردی مؤثر و مقیاس‌پذیر برای برآورد نفوذپذیری نسبی بر اساس داده‌های متداول مغزه‌گیری ارائه می‌دهد. بر اساس نتایج، مدل *Extra Trees* برای پیش‌بینی نفوذپذیری نسبی آب و مدل *KNN* برای پیش‌بینی نفوذپذیری نسبی نفت توصیه می‌شوند. به‌کارگیری این مدل‌ها می‌تواند سرعت و دقت شبیه‌سازی‌های مخزن را افزایش دهد، هزینه‌های عملیاتی را کاهش دهد و از تصمیم‌گیری‌های آگاهانه‌تر در توسعه میادین پشتیبانی کند.

۵. مراجع

۱. Muccino, J.C., W.G. Gray, and L.A.J.R.o.G. Ferrand, *Toward an improved understanding of multiphase flow in porous media*. 1998. **36**(3): p. 401-422.
۲. Chapter 7 *Relative permeability*, in *Developments in Petroleum Science*, J.R. Fanchi, Editor. 2000, Elsevier. p.

می‌توانند بینش‌های ارزشمندی در خصوص تغییرات نفوذپذیری نسبی تحت این شرایط ارائه دهند و طراحی و پایش فرایند ازدیاد برداشت را بهبود بخشند.

در مجموع، استفاده از یادگیری ماشین در برآورد نفوذپذیری نسبی منجر به شناسایی کارآمدتر مخزن، کاهش هزینه‌های پروژه و بهبود تصمیم‌گیری در کل چرخه عمر میدان می‌شود. با تداوم تحول دیجیتال در صنعت نفت و گاز، این ابزارهای هوشمند به تدریج به بخشی استاندارد از جعبه‌ابزار مهندسی مخزن تبدیل خواهند شد.

۳.۳. کاربردها و پیشنهادات آینده

نتایج این پژوهش نشان می‌دهد که مدل‌های یادگیری ماشین می‌توانند به‌عنوان روشی سریع، مقرون‌به‌صرفه و قابل اعتماد برای برآورد نفوذپذیری نسبی مورد استفاده قرار گیرند و جایگزین مناسبی برای آزمایش‌های آزمایشگاهی پرهزینه و زمان‌بر باشند. این مدل‌ها قابلیت به‌کارگیری در فرآیندهای مهندسی مخزن شامل شبیه‌سازی جریان چندفازی، طراحی و ارزیابی روش‌های ازدیاد برداشت، تحلیل عدم قطعیت و پایش میدانی را دارند و می‌توانند از طریق به‌روزرسانی مداوم با داده‌های جدید، در چرخه توسعه مخزن نقش پویایی ایفا کنند. با توجه به نتایج به‌دست‌آمده، توسعه این پژوهش در آینده می‌تواند شامل استفاده از پایگاه‌داده‌های گسترده‌تر تحت شرایط دما و فشار مخزن، مقایسه عملکرد مدل‌ها با شبکه‌های عصبی عمیق و مدل‌های ترکیبی، اعتبارسنجی میدانی، و ادغام رویکردهای داده‌محور با مدل‌سازی مبتنی بر فیزیک باشد. چنین توسعه‌ای می‌تواند منجر به ایجاد چارچوبی دقیق‌تر و قابل‌انتشار در کاربردهای صنعتی و تصمیم‌گیری‌های واقعی مهندسی مخزن گردد.

۴. نتیجه‌گیری

در این پژوهش، کاربرد چهار الگوریتم یادگیری ماشین شامل *Extra Trees*، *KNN*، *CatBoost* و *XGBoost* برای برآورد نفوذپذیری نسبی نفت و آب با استفاده از داده‌های مغزه از یکی از مخازن جنوب‌غرب ایران مورد بررسی قرار گرفت. یافته‌ها نشان دادند که مدل‌های مختلف برای پیش‌بینی نوع خاصی از نفوذپذیری عملکرد بهتری دارند و این امر بر ضرورت انتخاب مدل متناسب با نوع سیال تأکید می‌کند.

- and Exhibition? 2021. SPE.
۲۰. Zhao, B., et al., *A hybrid approach for the prediction of relative permeability using machine learning of experimental and numerical proxy SCAL data*. 2020. **25**(05): p. 2749-2764.
۲۱. Zhao, J., et al., *A Permeability Prediction Model of Single-Peak NMR T 2 Distribution in Tight Sandstones: A Case Study on the Huangliu Formation, Yinggehai Basin, China* : (۶)۵۶ . ۲۰۲۴ .p. 1303-1333.
۲۲. Okon, E.I and D. Appah. *Application of machine learning techniques in reservoir characterization*. in *SPE Nigeria Annual International Conference and Exhibition*. 2021. SPE.
۲۳. Dehghani, M.R., et al., *Estimation the pH of CO₂-saturated NaCl solutions using gene expression programming: Implications for CO₂ sequestration*. 2025. **25**: p. 104047.
۲۴. Cao, L., et al., *Interpretable Soft Sensors using Extremely Randomized Trees and SHAP*. 2023. **56**(2): p. 8000-8005.
۲۵. Okoro, E.E ., et al., *Application of artificial intelligence in predicting the dynamics of bottom hole pressure for under-balanced drilling: Extra tree compared with feed forward neural network model*. 2022. **8**(2): p. 227-236.
۲۶. Salih, A.K. and H.A.A.J.T.I.G.J. Hussein ,*Lost circulation prediction using decision tree, random forest, and extra trees algorithms for an Iraqi oil field*. 2022: p. 111-127.
۲۷. Dehghani, M.R., et al., *Estimation of hydrogen solubility in aqueous solutions using machine learning techniques for hydrogen storage in deep saline aquifers*. 2024. **14**(1): p. 25890.
۲۸. Omisakin, S. *Ensemble Machine Learning Methods to Predict Oil Production*. in *Innovations and Interdisciplinary Solutions for Underserved Areas: 7th International Conference, InterSol 20* , ۲۴Dakar, Senegal, July 3–4, 2024, *Proceedings*. 2025. Springer Nature.
۲۹. Hashemizadeh, A., et al., *Experimental measurement and modeling of water-based drilling mud density using adaptive boosting decision tree, support vector machine, and K-nearest neighbors: A case study from the South Pars gas field*. 2021. **207**: p. 109132.
۳۰. Khamis, Y.E., et al., *Rate of penetration prediction in drilling operation in oil and gas wells by k-nearest neighbors and multi-layer perceptron algorithms*. 2023. **14**(3): p. 75. ۷۷۰-۵
۳۱. Martín-Martín, M., et al., *Using python libraries and k-Nearest neighbors algorithms to delineate syn-sedimentary faults in sedimentary porous media*. 2023. **153**: p. 106283.
۳۲. Saddiqi, H.A., et al., *Modelling and predicting lift force and trans-membrane pressure using linear, KNN, ANN and response surface models during the separation of oil drops from produced water*. 2024. **66**: p. 106014.
۳۳. Mohammadi, M.-R., et al., *Modeling hydrogen solubility in hydrocarbons using extreme gradient boosting and equations of state*. 2021. **11**(1): p. 17911.
۳۴. Yakoot, M.S.e., A.M.S. Ragab, and O. Mahmoud. *Multi-* 85-101.
۳. Viera, M.A.D., et al., *Mathematical and numerical modeling in porous media: Applications in geosciences*. 2012: CRC Press.
۴. Honarpour, M.M., G.V. Chilingarian, and S.J. Mazzullo, *Chapter 8 Permeability and Relative Permeability of Carbonate Reservoirs*, in *Developments in Petroleum Science*, G.V. Chilingarian, S.J. Mazzullo, and H.H. Rieke, Editors. 1992, Elsevier. p. 399-416.
۵. Porges, F., *Chapter 5. Relative Permeability Concepts*. 2006.
۶. Honarpour, M. and S.J.J.o.p.t. Mahmood, *Relative-permeability measurements: An overview*. 1988. **40**(08): p. 963-966.
۷. Siginer, D. and S.J.J.A.M. Bakhtiyarov, *Flow in porous media of variable permeability and novel effects*. 2001. **68**(2): p. 312-319.
۸. Malkovsky, V., A. Zharikov, and V.J.I. Shmonov, *Physics of the Solid Earth, New methods for measuring the permeability of rock samples for a single-phase fluid*. 2009. **45**: p. 89-100.
۹. Bryant, S. and M.J.P.r.A. Blunt, *Prediction of relative permeability in simple porous media*. 1992. **46**(4): p. 2004.
۱۰. Schembre, J., A.J.J.o.P.S. Kovscek, and Engineering, *A technique for measuring two-phase relative permeability in porous media via X-ray CT measurements*. 2003. **39**(1-2): p. 159-174.
۱۱. Alizadeh, A. and M.J.R.o.G. Piri, *Three-phase flow in porous media: A review of experimental studies on relative permeability*. 2014. **52**(3): p. 468-521.
۱۲. Sigmund, P. and F.J.S.o.P.E.J. McCaffery, *An improved unsteady-state procedure for determining the relative-permeability characteristics of heterogeneous porous media (includes associated papers 8028 and 8777)*. 1979. **19**(01): p. 15-28.
۱۳. Hao, L., P.J.I.J.o.H. Cheng, and M. Transfer, *Pore-scale simulations on relative permeabilities of porous media by lattice Boltzmann method*. 2010. **53**(9-10): p. 1908-1913.
۱۴. Blom, S ., J. Hagoort, and D.J.S.J. Soetekouw, *Relative permeability at near-critical conditions*. 2000. **5**(02): p. 172-181.
۱۵. Hossain, Z. *Relative permeability prediction from Image Analysis of thin sections*. in *SPE Europec featured at EAGE Conference and Exhibition?* 2011. SPE.
۱۶. Ecay, L., et al., *On the prediction of permeability and relative permeability from pore size distributions*. 2020. **133**: p. 106074.
۱۷. Kadet, V., A.J.J.o.P.S. Galechyan, and Engineering, *Percolation modeling of relative permeability hysteresis*. 2014. **119**: p. 139-148.
۱۸. Wang, Y., et al. *The application of machine learning algorithm in relative permeability upscaling for oil-water system*. in *International Petroleum Technology Conference*. 2021. IPTC.
۱۹. Alfonso, C.E., F. Fournier, and V .Alcobia. *A machine learning methodology for rock-typing using relative permeability curves*. in *SPE Annual Technical Conference*

- class taxonomy of well integrity anomalies applying inductive learning algorithms: analytical approach for artificial-lift wells. in SPE Annual Technical Conference and Exhibition? 2021. SPE.*
۳۵. Shakouri, S. and M.J.E. Mohammadzadeh-Shirazi, *Modeling of asphaltic sludge formation during acidizing process of oil well reservoir using machine learning methods. 2023. 285: p. 129433.*
۳۶. Shakouri, S. and M.J.P.S. Mohammadzadeh-Shirazi, *Machine learning approaches for assessing stability in acid-crude oil emulsions: Application to mitigate formation da. 2025. 22.(۲)*
۳۷. Xie, Z., et al., *Prediction of conformance control performance for cyclic-steam-stimulated horizontal well using the XGBoost: a case study in the Chunfeng heavy oil reservoir. 2021. 14(23): p. 8161.*
۳۸. Zarin, T., H.A. Eshkaftaki, and A.J.R.i.E. Sharifi, *Machine Learning-Based Prediction of Oil-Water Relative Permeability Using Core Flooding and CT-Scan Data. 2025: p. 105735.*
۳۹. Alali, R., K. Abe, and H.J.A.o. Fujii, *Simulation of Core Flooding with Predicted Oil and Water Relative Permeabilities Using Bagging, Boosting, and Stacking Machine Learning Techniques. 2025. 10(16): p. 15967-15978.*