

## تعیین واحدهای جریان هیدرولیکی با استفاده از روش جنگل تصادفی برای یکی از مخازن نفتی ایران

مهدی رستگاری\*، شرکت پتروپارس زاگرس • علی صنعتی، دانشگاه حکیم سبزواری

چکیده

اطلاعات مقاله

تاریخ ارسال نویسنده: ۹۵/۰۱/۲۳

تاریخ ارسال به داور: ۹۵/۰۱/۲۶

تاریخ پذیرش داور: ۹۵/۰۳/۲۰

شبه‌سازی مخازن به دلیل ناهمگونی سنگ مخزن یکی از پیچیده‌ترین فعالیت‌ها در مهندسی نفت است. منظور از ناهمگونی، توزیع نامنظم اجزای تشکیل‌دهنده سنگ در کنار یکدیگر است که موجب می‌شود رفتار سنگ مخزن به‌درستی قابل پیش‌بینی نباشد. نظریه واحد جریان سیالات به منظور غلبه بر این سردرگمی بیان شده است به گونه‌ای که در این نظریه واحدهای هیدرولیکی مرتبط با خصوصیات زمین‌شناسی، جریان سیال در محیط متخلخل را کنترل می‌کنند. مهمترین پارامتری که بر جریان سیال در این محیط تأثیرگذار است، ویژگی هندسی گلوگاه‌های موجود در منافذ سنگ می‌باشد که این ویژگی‌های هندسی به نوبه خود تحت تأثیر عواملی همچون کانی‌شناسی (نوع، فراوانی و محل قرارگیری) و بافت سنگ (اندازه دانه، شکل دانه و جورشدگی) قرار دارند. واحدهای جریان هیدرولیکی عموماً از روی داده‌های مغزه و در آزمایشگاه به‌دست می‌آیند که به‌واقع کار دشوار و زمان‌بری می‌باشد. بنابراین به‌کارگیری روش‌هایی که از نگاره‌های موجود از چاه برای به‌دست آوردن واحدهای جریان هیدرولیکی بهره می‌گیرند بسیار مورد توجه قرار گرفته‌اند. یکی از این روش‌ها روش جنگل تصادفی است. هدف از ارائه این مقاله، استفاده موردی از الگوریتم جنگل تصادفی برای تعیین واحدهای جریان هیدرولیکی از روی نمودارهای چاه‌نگاری در چاه‌هایی است که فاقد داده‌های مغزه هستند. نتایج به‌دست آمده از این مقاله نشان می‌دهد روش جنگل تصادفی به‌خوبی قابلیت تعیین واحدهای جریان هیدرولیکی در چاه‌های جنوب‌غرب ایران را با دقت ۸۱ درصد داراست. همچنین این نتایج با نتایج ارزیابی پتروفیزیکی و تراوایی حاصل از ابزار DSI مقایسه و مشاهده شده است که نتایج به‌دست آمده از تطابق بالایی با یکدیگر برخوردارند. بنابراین روش جنگل تصادفی می‌تواند به‌عنوان روشی کم‌هزینه و موثر در مقابل آنالیز آزمایشگاهی و پرهزینه مغزه در صنعت مورد استفاده قرار گیرد.

واژگان کلیدی:

روش جنگل تصادفی، واحدهای جریان هیدرولیکی، مخازن نفتی، نمودار صوتی برشی دو قطبی

مقدمه

کردند [۱، ۲]. آنها با ترکیب قانون داری برای حرکت سیال در محیط متخلخل و قانون پویزلی برای حرکت سیال در لوله‌ها، رابطه‌ای را بین تخلخل و تراوایی ارائه دادند. رابطه به‌دست آمده توسط کوزنی و کارمن معیار مناسبی برای زون‌بندی مخزن نیست، زیرا برای هر تخلخل مشخص در یک نوع سنگ، تراوایی می‌تواند در مقادیر مختلفی تغییر کند که نشان‌دهنده وجود واحدهای جریان هیدرولیکی متفاوت است. تئوری واحدهای جریان هیدرولیکی اساساً توسط آمیفل بیان شده است. توسعه کاربرد این روش به علت مسئله تخمین تراوایی در چاه‌هایی است که اطلاعات مغزه در آنها موجود نیست ولی دارای نگاره‌های چاه‌پیمایی هستند [۳]. عباس‌زاده و همکاران برخی از روش‌های دسته‌بندی که منجر به تقسیم‌بندی واحدهای جریان هیدرولیکی می‌شود را مورد بررسی قرار دادند که از آنجمله می‌توان به ارزیابی هیستوگرام، ارزیابی گراف و الگوریتم تحلیلی اشاره کرد [۴]. فهاد و همکاران با استفاده از

برای محاسبه واحدهای جریان هیدرولیکی به‌طور متداول از داده‌های مغزه (تخلخل موثر و تراوایی) استفاده می‌گردد. از آنجایی که تعیین این واحدها ارتباط مستقیمی با تراوایی دارد، پارامترهای موثر بر جریان سیال که عمدتاً شامل شکل هندسی حفرات و ارتباط گلوگاه‌ها هستند، بسیار موثر می‌باشند. شکل هندسی حفرات به نوبه خود توسط کانی‌شناسی (نوع، فراوانی و موقعیت) و بافت سنگ (اندازه دانه، شکل دانه، جورشدگی و نحوه قرارگیری ذرات کنار یکدیگر) کنترل می‌شود. ترکیبات متفاوتی از این خصوصیات زمین‌شناسی می‌تواند منجر به ایجاد واحدهای جریانی مجزایی گردد که هر یک دارای خواص منحصر به فرد گذردهی سیال هستند. در حقیقت گروه‌بندی سنگ‌ها بر مبنای مشخصه‌های بنیادی جریان، اساس طبقه‌بندی واحدهای هیدرولیکی است.

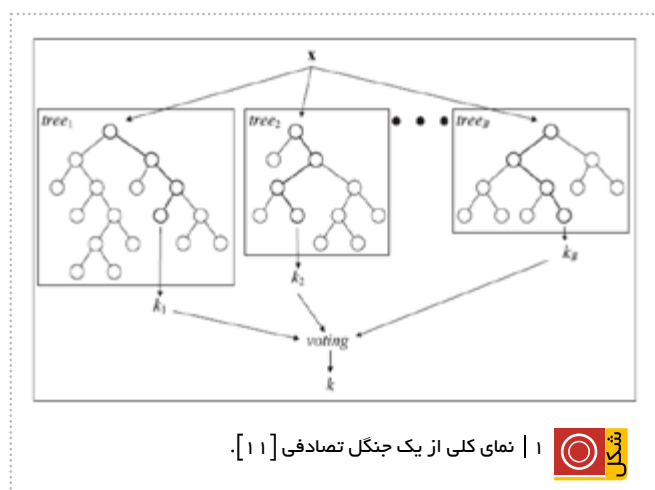
کوزنی و کارمن برای یافتن ارتباطی مناسب بین تخلخل و تراوایی، محیط متخلخل را به‌صورت مجموعه‌ای از لوله‌های مؤئنه مدل‌سازی

\* نویسنده‌ی عهده‌دار مکاتبات (mahdi65.rastgarnia@gmail.com)

۲) تعداد  $m$  متغیر به تصادف از بین  $M$  متغیر توضیحی انتخاب می‌گردد ( $m < M$ ). معمولاً تعداد متغیرها با  $m_{try}$  نشان داده می‌شود که بایستی توسط کاربر انتخاب گردد. پیشنهاد شده است که در مدل رگرسیونی،  $m = \frac{M}{3}$  و در مدل کلاس بندی  $m = \sqrt{M}$  در نظر گرفته شود [۹]. وقتی که درخت ساخته شد، داده OOB در درخت ساخته شده قرار داده می‌شود و میزان خطا محاسبه می‌گردد. از آنجایی که تعداد درخت‌های ساخته شده بر روی نمونه‌های Boot-strap زیاد است این میزان خطا ناریب می‌شود. از داده‌های OOB برای برآورد ناریب خطا و برآورد متغیر بااهمیت استفاده می‌گردد. در واقع، کلاسه‌بندی روی درخت‌های ساخته شده روی OOB ها انجام شده است. معیاری که در این قسمت برای میزان دقت در OOB استفاده می‌شود خطای Misclassification است [۱۰].

در شکل-۱ نمای کلی از الگوریتم جنگل تصادفی نشان داده شده است. با افزایش تعداد درخت‌ها اثر بیش برآزش که در روش درخت تصمیم رخ می‌دهد، از بین می‌رود. بدین صورت که در هر مرحله از رشد درخت، مدل دقیق‌تر شده و  $\text{sum of square}$  کاهش می‌یابد. یک راه برای توقف رشد درخت می‌تواند بر اساس میزان کاهش Sum of Square باشد، بدین ترتیب که اگر در یک مرحله از رشد درخت،  $\text{sum of square}$  چهار کاهش چندانی نشود می‌توان از رشد درخت در آن مرحله صرف‌نظر نمود. حال این میزان کاهش در  $\text{sum of square}$  می‌تواند در دست کاربر باشد. هدف در این روش این است که کلاس‌بندی به گونه‌ای انجام شود که دارای مینیم مقدار خطا باشد. در این روش مقدار بهینه برای دو پارامتر تعداد درخت تصمیم و تعداد متغیرهایی که در ساخت درخت تصمیم به کار می‌رود، مورد بررسی قرار می‌گیرد. به عبارتی دیگر، مینیم کردن خطا در هر درخت تصمیم، معادل با مقدار بهینه برای  $n_{tree}$  و  $m_{try}$  است.

همچنین با توجه به شکل-۱، اگر یک جنگل تصادفی ساخته شده باشد، با ورود یک بردار  $x$  به داخل جنگل تصادفی، این بردار وارد تمامی درخت‌ها می‌شود و در نودهای پایانی قرار می‌گیرد.



شکل ۱ | نمای کلی از یک جنگل تصادفی [۱۱].

روش واحدهای جریان هیدرولیکی، تراوایی را از داده‌های نگارهای چاه‌پیمایی در یک مخزن ماسه‌سنگی عربستان تخمین زده‌اند [۵]. پراسد به بهبود ارتباط بین سرعت امواج و تراوایی درون واحدهای جریان هیدرولیکی اشاره نموده است [۶] و در نهایت کاظم‌زاده و همکاران ارتباط مقاومت ویژه سازندی و تخلخل را در واحدهای جریان هیدرولیکی بررسی نموده‌اند [۷].

در این مطالعه با استفاده از واحدهای جریان هیدرولیکی مربوط به چاه-۱، این واحدهای جریان هیدرولیکی برای چاه‌های-۲ و ۳ با روش جنگل تصادفی و با بهره‌گیری از نگارهای متداول تخمین زده شده‌اند. این مطالعه در قسمت پایین مخزن آسماری که بیشتر ماسه‌سنگی است، انجام گرفته است.

### ۱-روش کار

در این تحقیق، روش جنگل تصادفی با استفاده از داده‌های چاه-۱ آموزش داده می‌شود. این روش در نرم‌افزار R اجرا شده است. R، یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای محاسبات آماری و تحلیل داده است که حاوی محدوده گسترده‌ای از تکنیک‌های آماری (از جمله: مدل‌سازی خطی و غیرخطی، آزمون‌های کلاسیک آماری، تحلیل سری‌های زمانی، رده‌بندی، خوشه‌بندی و غیره) و قابلیت‌های گرافیکی می‌باشد.

### ۱-۱- روش جنگل تصادفی

اساس کار روش جنگل تصادفی (RF)، ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه‌ی بوت استرپ از داده‌ها شرکت دارد و در ساخت هر درخت به‌طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. روش بوت استرپ، روش نمونه‌گیری با جای‌گذاری است. لذا با تکرار عملیات نمونه‌گیری، تعدادی مجموعه داده OOB از مجموعه آموزشی به‌وجود می‌آید که می‌توان برای هر مجموعه داده‌ی آموزشی یک درخت تصمیم محاسبه نمود [۸].

الگوریتم کلی مدل RF به این صورت است: فرض کنید تعداد داده‌هایی که به عنوان training data در دست است و در ساخت مدل حضور دارند،  $N$  باشد. همچنین تعداد کل متغیرهای پیش‌بینی‌کننده موجود را با  $M$  و تعداد کل درخت‌های حاضر در مدل را  $n$  ( $n_{tree}$ ) در نظر بگیرد. برای ساخت درخت  $i$ ام ( $i=1, \dots, n_{tree}$ ) کافی است الگوریتم زیر دنبال گردد:

۱) در این روش، نمونه اصلی به دو قسمت نمونه آموزشی و نمونه امتحانی تقسیم می‌شود. سپس از نمونه آموزشی نمونه‌گیری مجددی به حجم  $N$  گرفته می‌شود و  $1/3$  از نمونه آموزشی جدید، به عنوان نمونه خارج از کیسه (OOB) از مجموعه داده‌های آموزشی جدا می‌گردد. در واقع OOB، در هر درخت، به نوعی نقش نمونه امتحانی (test set) را برای آن درخت ایفا می‌کند.

میانگین کاهش دقت است. در این مطالعه متغیر بااهمیت با استفاده از میانگین کاهش ضریب جینی به کار برده شده است. در واقع، ضریب جینی نوعی خطاست که رابطه آن در زیر آمده است.

$$Gini - index = \sum_{k=1}^K p_{mk} \hat{p}_{mk} (1 - \hat{p}_{mk}) \quad (1)$$

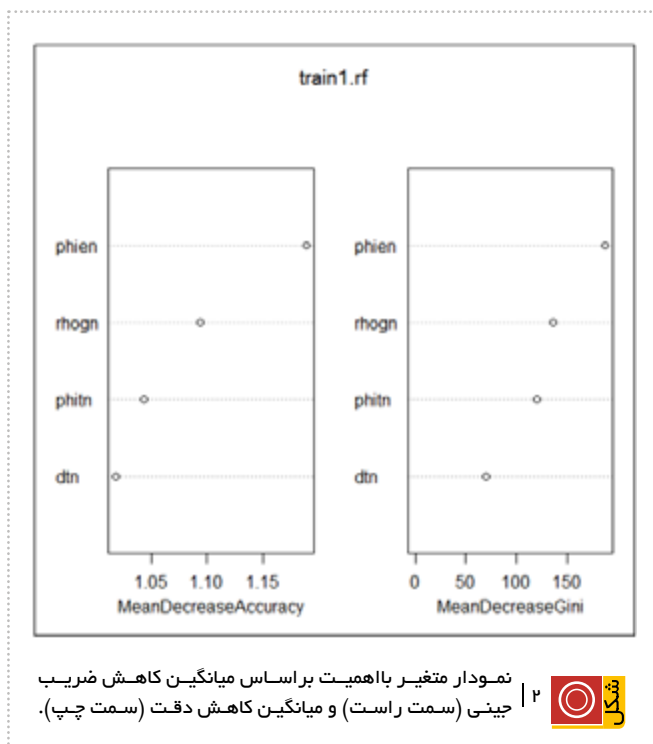
در فرمول بالا  $\hat{P}_m$  مقدار احتمال درستی کلاس‌بندی را نشان می‌دهد. ضریب جینی از حاصل ضرب احتمال درست و نادرست بودن کلاس‌بندی نتیجه می‌شود [۱۲].

شکل ۲- نمودار متغیر بااهمیت را برای دقت و ضریب جینی داده‌های آموزشی نشان می‌دهد که منظور از دقت، میزان درستی متغیر بااهمیت می‌باشد در حالی که ضریب جینی نوعی خطاست. به عبارت دیگر، این دو معیار مخالف هم هستند.

شکل ۲- نمودار متغیر بااهمیت براساس میانگین کاهش ضریب جینی (سمت راست) و میانگین کاهش دقت (سمت چپ). در هر کدام از نمودارهای شکل ۲- متغیرها به صورت نزولی مرتب شده‌اند. در حالتی که معیار مورد نظر ضریب جینی است متغیرهایی که در پایین نمودار هستند، دارای خطای کمتر و اهمیت بیشتری می‌باشند. همچنین در نمودار دقت، متغیرهایی که دارای اهمیت بیشتری هستند در پائین قرار گرفته‌اند. بدین ترتیب متغیرها بر اساس اهمیتشان در مدل مرتب شده‌اند.

## ۲- بحث و بررسی

تخمین واحدهای جریان هیدرولیکی با استفاده از روش آنالیز مغزه



سپس، بردار ورودی  $x$  در هر کدام از درخت‌ها کلاسی را به خود اختصاص می‌دهد. در نهایت، از بین تمامی کلاس‌ها آن کلاسی که بیشترین تعداد (آراء) را داشته باشد، کلاس پیش‌بینی جنگل تصادفی برای بردار ورودی  $x$  خواهد بود [۹].

روش جنگل تصادفی در نرم افزار R دارای تابعی با دو پارامتر  $mtry$  و  $ntree$  به ترتیب تعداد متغیرها و تعداد درخت می‌باشد که قابل تغییر بوده و توسط کاربر تعیین می‌شود. کلاسه‌بندی جنگل تصادفی ابتدا بر روی داده آموزشی و سپس روی داده‌های امتحانی اجرا می‌گردد. برای این منظور باید تابعی در نظر گرفته شود و سپس برای دو پارامتر  $mtry$  و  $ntree$  مقادیری به صورت اختیاری فرض گردد. قابل به ذکر است که برای کلاسه‌بندی به صورت پیش فرض  $M$ : تعداد کل متغیر) پیشنهاد می‌گردد [۹]. در نهایت مدلی که دارای کمترین خطای OOB باشد، انتخاب می‌گردد.

در این مطالعه ابتدا داده‌های چاه-۱ به دو قسمت آموزشی ( $train$ ) و امتحانی ( $test$ ) تقسیم گردیده و سپس تابع مورد نظر بر روی داده‌ها اجرا شده است. شکل ۱- برنامه استفاده شده در این تحقیق را نشان می‌دهد. جدول ۱- بیانگر مقادیر بهینه حاصل شده از اجرای این برنامه برای پارامترهای  $mtry$  و  $ntree$  است.

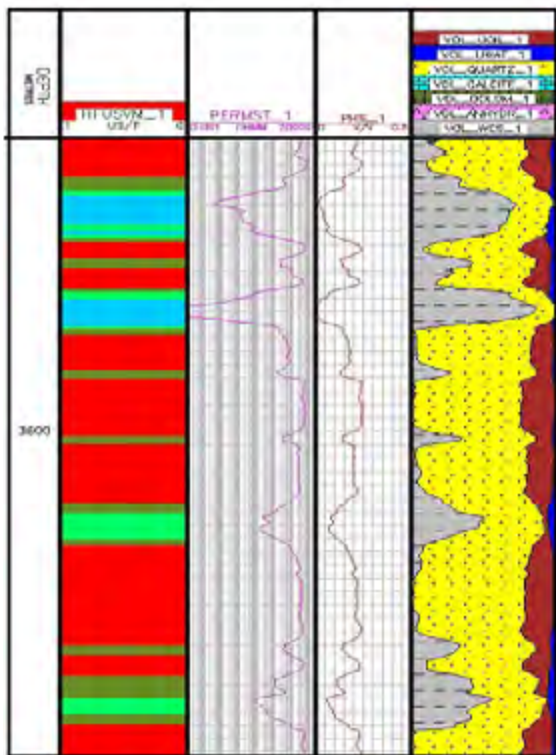
جدول ۱ | مقادیر دقت برای  $mtry$  و  $ntree$  های متفاوت.

Accuracy (دقت)	ntree	Mtry
۰/۸۰۹۸	۵۰۰	۲
۰/۸۱۳۳	۱۰۰۰	۲
۰/۸۰۸۶	۱۵۰۰	۲
۰/۸۰۵۱	۵۰۰	۳
۰/۸۰۴	۱۰۰۰	۳
۰/۸۰۳۸	۱۵۰۰	۳
۰/۴۷	۰/۲۳	۰/۲

### ۲-۱- انتخاب متغیر در جنگل تصادفی

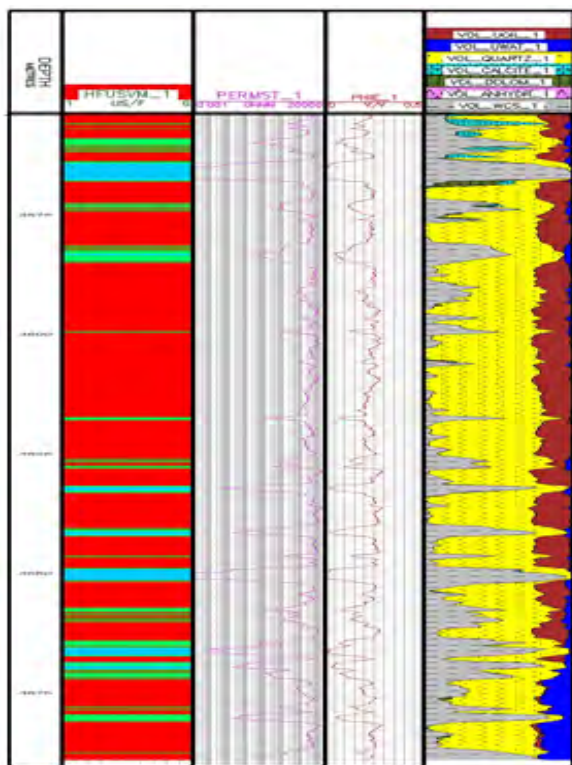
مسئله بعدی که در روش جنگل تصادفی اهمیت پیدا می‌کند، انتخاب متغیر است. زیرا تمامی متغیرها برای مدل خوب نبوده و بعضی از آنها زائد و اضافی می‌باشند. در واقع، این کار منجر به کاهش ابعاد داده‌ها، محاسبات کمتر و هزینه پائین می‌شود. انتخاب متغیر به مفهوم تحلیل قوی در کلاس‌بندی است. این روش انتخاب زیرمجموعه‌ای از فضای اصلی داده‌هاست که معادل با بهبود یافتن کیفیت داده‌ها می‌باشد [۱۱].

محاسبه متغیر بااهمیت بر اساس میانگین کاهش ضریب جینی و

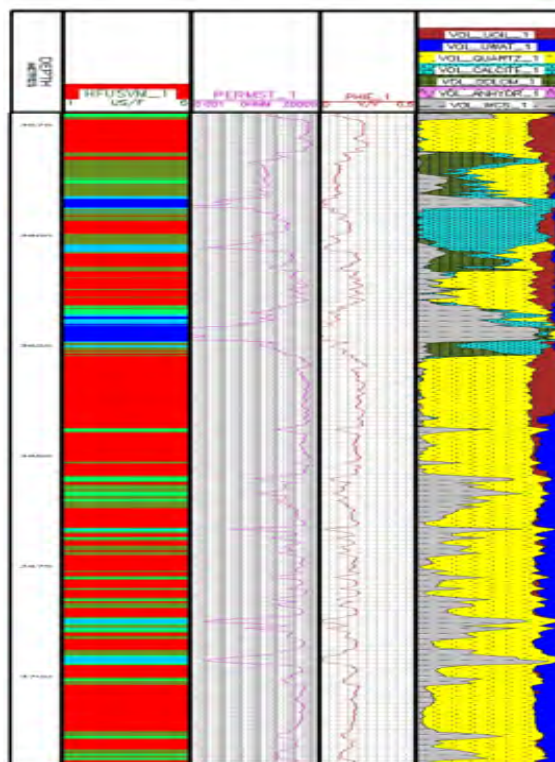


شکل ۴ | واحدهای جریان هیدرولیکی تخمین زده شده (ستون آخر از سمت چپ) برای چاه-۲

به صورت گسسته و با هزینه‌های فراوان انجام می‌گیرد در حالی که با استفاده از روش جنگل تصادفی، این پارامتر به صورت پیوسته برای کل بازه‌ی چاه با هزینه بسیار پایین و با دقت بسیار بالا در چاه‌های فاقد مغزه تخمین زده می‌شود. در این مطالعه، تطابق بالای نتایج حاصل از تراوایی حاصل شده از ابزار DSI و ارزیابی پتروفیزیکی با نتایج روش جنگل تصادفی موید این مطلب است. مدل ایجاد شده در این مطالعه برای تخمین واحدهای جریان هیدرولیکی چاه‌های ۲ و ۳ به کار گرفته شده که فقط داده‌های ورودی (تخلخل موثر:  $\phi_{\text{eff}}$ ، صوتی:  $\text{dtn}$ ، تخلخل کل:  $\phi_{\text{itn}}$ ، چگالی:  $\rho_{\text{hgn}}$ ) را شامل می‌گردد و توسط آن واحدهای جریان هیدرولیکی مختلف برای دو چاه مذکور به دست آمده است. در شکل ۳، چاه-۱ که برای ایجاد مدل‌سازی روش جنگل تصادفی به کار برده شده، نشان داده شده است. ستون اول از سمت راست حجم نفت (به رنگ قهوه‌ای)، حجم آب (به رنگ آبی) و لیتولوژی سازند ارزیابی شده توسط روش‌های پتروفیزیکی، ستون دوم تخلخل موثر، ستون سوم تراوایی حاصل از ابزار DSI و ستون چهارم واحدهای جریان هیدرولیکی را نشان می‌دهد. واحدهای جریان هیدرولیکی از لحاظ بهترین رخساره مخزنی رتبه‌بندی شده‌اند که واحدهای جریان به رنگ‌های قرمز، سبز لجنی، سبز روشن، آبی آسمانی و آبی تیره به ترتیب رتبه ۱ تا ۵ را به خود اختصاص داده‌اند. شکل ۴ و ۵ واحدهای جریان هیدرولیکی



شکل ۵ | واحدهای جریان هیدرولیکی تخمین زده شده (ستون آخر از سمت چپ) برای چاه-۳



شکل ۳ | چاه-۱ برای ایجاد مدل‌سازی روش جنگل تصادفی

در پایان لازم به ذکر است که روش جنگل تصادفی برای حجم داده‌های زیاد به خوبی عمل می‌کند و از دقت بالایی برخوردار است. در این روش، هر درخت به طور کامل رشد می‌کند و هرس نمی‌شود و این ویژگی باعث می‌شود که مدل نهایی دچار بیش-برآورد نگردد. بدین ترتیب که اگر در یک مرحله از رشد درخت، sum of square دچار کاهش چندانی نشود، می‌توان از رشد درخت در آن مرحله صرف نظر کرد. ■

جدول ۲ | مشخصات واحدهای جریان معرفتی شده

نام رخساره	تراوایی (mD)	تخلخل (%)
رخساره-۵	۲۰۳۰	۱۹
رخساره-۴	۴۶	۱۲
رخساره-۳	۰/۹	۸
رخساره-۲	۰/۰۱	۲/۳
رخساره-۱	۰/۰۰۲	۱

تخمین زده شده را توسط روش جنگل تصادفی نشان می‌دهد.

### نتیجه گیری

همان طور که در شکل‌های ۴- و ۵ مشاهده می‌شود، واحدهای جریان هیدرولیکی به دست آمده از روش جنگل تصادفی تطابق بسیار خوبی را با حجم نفت (ستون اول از سمت راست) و تخلخل موثر ارزیابی شده از روش‌های پتروفیزیکی (ستون دوم از سمت راست) و همچنین با نتایج تراوایی حاصل شده از ابزار DSI (ستون سوم از سمت راست) نشان می‌دهد. جدول ۲- واحدهای جریان معرفتی شده، همراه با مشخصات ارائه نموده است. در این مطالعه با استفاده از روش جنگل تصادفی، واحدهای جریان هیدرولیکی برای چاه‌های فاقد مغزه، با دقت ۸۱ درصد تخمین زده شده است. روش جنگل تصادفی تعداد بسیار زیادی از متغیرهای ورودی را می‌تواند در بر بگیرد بدون اینکه تغییری را حذف نماید، به این صورت که این روش برآوردی برای اهمیت متغیرها که در کلاس بندی مهم هستند ارائه می‌دهد. در این مطالعه نگار صوتی (dtn) با اهمیت ترین متغیر ورودی به دست آمده است.

### پانویس ها

1. Random Forest
2. Bootstrap
3. Out Of Bag
4. Misclassification
5. ntree
6. mtry
7. Variable Importance

### منابع

- [1] Carman, P.C., Fluid flow through granular beds. Transactions-Institution of Chemical Engineers, 1937. 15: p. 150-166.
- [2] Kozeny, J., Uber kapillare Leitung des Wassers in Boden, Sitzungsberichte. Royal Academy of Sciences, Vienna. Proc. Class I, 1927. 136.
- [3] Amaefule, J.O., et al. Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells. in SPE Annual Technical Conference and Exhibition. 1993. Society of Petroleum Engineers.
- [4] Abbaszadeh, M., H. Fujii, and F. Fujimoto, Permeability prediction by hydraulic flow units-theory and applications. SPE Formation Evaluation, 1996. 11(04): p. 263-271.
- [5] Al-Ajmi, F.A. and S.A. Holditch. Permeability estimation using hydraulic flow units in a Central Arabia reservoir. in SPE Annual Technical Conference and Exhibition. 2000. Society of Petroleum Engineers.
- [6] Prasad, M., Velocity-permeability relations within hydraulic units. Geophysics, 2003. 68(1): p. 108-117.
- [7] Kazemzadeh, E., B.H.M. NABI, and M.R. REZAEI, The study of formation resistivity factor by using hydraulic flow units method in carbonate reservoirs. 2008.
- [8] Liaw, A. and M. Wiener, Classification and regression by randomForest. R news, 2002. 2(3): p. 18-22.
- [9] Breiman, L., Random forests. Machine learning, 2001. 45(1): p. 5-32.
- [10] Verikas, A., A. Gelzinis, and M. Bacauskiene, Mining data with random forests: A survey and results of new tests. Pattern Recognition, 2011. 44(2): p. 330-349.
- [11] Genuer, R., J.-M. Poggi, and C. Tuleau-Malot, Variable selection using random forests. Pattern Recognition Letters, 2010. 31(14): p. 2225-2236.
- [12] Hastie, T., R. Tibshirani, and J. Friedman, The elements of statistical learning 2nd edition. 2009, New York: Springer.