

مقاله پژوهشی

طراحی مدل گروهی تخمین تراوایی مخزن هیدروکربوری با استفاده از نگاره‌های پتروفیزیکی بر اساس تفکیک لیتولوژیکی

عباس سلحشور^{۱*}؛ احمد گابینی^۲؛ علیرضا شاهین^۳؛ مصیب کمری^۴

۱. دانشجوی دکترای تخصصی؛ مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه ایوانکی

۲. استادیار؛ مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه ایوانکی

۳. دکترای تخصصی؛ زمین‌شناسی، دانشکده علوم، دانشگاه اصفهان

۴. کارشناسی ارشد؛ مهندس ارشد پتروفیزیک، شرکت ملی مناطق نفت خیز جنوب

دریافت مقاله: ۱۴۰۰/۰۵/۰۷ پذیرش مقاله: ۱۴۰۰/۱۰/۲۲

شناسه دیجیتال (DOI): 10.22107/jpg.2022.297342.1148

چکیده

واژگان کلیدی

تراوایی یا نفوذپذیری، یکی از خصوصیات مهم مخازن نفت و گاز است که پیش‌بینی آن دشوار می‌باشد. در حال حاضر از مدل‌های تجربی و رگرسیونی برای پیش‌بینی تراوایی استفاده می‌شود، از سوی دیگر افزایش دقت در پیش‌بینی تراوایی جهت نقاطی که فاقد نمونه مغزه است از اهمیت ویژه‌ای در تحلیل رفتار مخزن برخوردار است. در چند وقت اخیر، به دلیل قابلیت پیش‌بینی بهتر، از الگوریتم‌های یادگیری ماشین برای پیش‌بینی تراوایی استفاده شده است. در این مطالعه، مدل یادگیری ماشین گروهی جدیدی برای پیش‌بینی تراوایی در مخازن نفت و گاز معرفی شده است. در این روش، داده‌های ورودی با استفاده از اطلاعات لیتولوژی لاگ‌ها برچسب‌گذاری شده و به تعدادی از خوشه‌ها تفکیک می‌شوند و هر خوشه توسط الگوریتم یادگیری ماشین مدل‌سازی شد. برخلاف مطالعات قبلی که به صورت مستقل روی مدل‌ها کار می‌کردند در اینجا ما ضمن طراحی یک مدل گروهی با استفاده از الگوریتم‌های رگرسیونی درخت تصمیم افزوده (ETR)، رگرسیون درخت تصمیم (DTR) و رگرسیون گرادیان تقویت شده (GBR) و داده‌های پتروفیزیکی، توانستیم صحت و دقت پیش‌بینی همچنین خطای میانگین مربعات را به طرز چشم‌گیری بهبود بخشیم و تراوایی را با دقت ۹۹.۸۲ درصد پیش‌بینی کنیم. نتایج نشان داد که مدل‌های گروهی در بهبود دقت پیش‌بینی تراوایی در مقایسه با مدل‌های انفرادی تأثیر فراوانی دارند و همچنین تفکیک نمونه‌ها بر اساس اطلاعات لیتوژی، دلیلی بر بهینه نمودن تخمین تراوایی نسبت به تحقیقات گذشته بود.

۱. پیش‌گفتار

رگرسیون خطی بین این دو متغیر و تعمیم آن به کل چاه است که چندان مناسب نیست. یکی از مدل‌های ارائه شده برای تخمین تراوایی، مدل کوزنی-کارمن است، مدلی کوزنی-کارمن از ادغام دو قانون داری، برای جریان سیال در محیط متخلخل و مدل پوزای، برای جریان در لوله‌ها، به دست آمده است (Abbaszadeh et al., 1996). محاسبه تراوایی با استفاده از مدل کوزنی-کارمن، به علت وجود پارامترهای پیچیدگی مسیر جریان، فاکتور شکل و نسبت سطح حفره به حجم دانه، مشکل بوده از این رو محققان با استفاده از روابط تخلخل-تراوایی در چارچوب واحدهای جریانی هیدرولیکی به منظور تعیین سنگ‌های مخزنی ناهمگن جهت

تراوایی توانایی سنگ‌های متخلخل در انتقال مایع است. کاربرد این متغیر مهم در مطالعات مخازن زیرزمینی نفت و گاز جهت مدیریت مخزن و فرآیندهای توسعه‌ی میدان‌های نفت و گاز، اجتناب‌ناپذیر است (Ramzi, 1998) (North, 1985). از لحاظ ماهیت و مشکلات عملیاتی، تخمین این متغیر نسبت به برخی متغیرهای دیگر سنگ و سیال، دشوارتر است (Oberto, 1984). شیوه‌هایی که در حال حاضر معمول است تعیین تراوایی و تخلخل مغزه در آزمایشگاه، به دست آوردن یک

بررسی دقیق رفتار مخزن و سیالات موجود در آن بهره‌برداری کردند. اولین مرحله پیش‌بینی رفتار مخزن طی تولید، تعیین گروه‌های سنگی مخزن است. با استفاده از گروه‌های سنگی مخزن می‌توان بخش‌های مخزنی را از بخش‌های غیر مخزنی تفکیک کرد (ایزدی، ۱۳۹۱). تلاش‌های مختلفی جهت تخمین تراوایی صورت گرفته که می‌توان آن‌ها را به دو دسته‌ی روش‌های تجربی و آماری تقسیم کرد. روش‌های آماری، روش‌های نوینی مبتنی بر هوش مصنوعی، شبکه‌های عصبی، منطق فازی و غیره هستند که بسیار کارآمد ظاهر شده‌اند. در این مقایسه باید تعداد متغیرهای مخزن مدنظر، کیفیت و حجم و تنوع اطلاعات آزمایشی انتخاب‌شده، همگن یا ناهمگن بودن مخزن که همگی بر دقت روش تأثیرگذار هستند را در نظر گرفت. در این مقاله سعی شده با استفاده از حداکثر متغیرهای در دسترس مرتبط با تراوایی (شامل نمودارهای پتروفیزیکی، طیف پرتو گاما، تخلخل، نوترون، چگالی سازند، صوتی، عامل فتوالکتریک، داده‌های تخلخل و درصد اشباع آب مغزه) به-عنوان ورودی و همچنین برچسب‌گذاری این مقادیر بر اساس اطلاعات لیتولوژی این نمودارها و تفکیک نمونه‌های مشابه در دسته‌های جداگانه، ضمن استفاده از ترکیب گروهی از الگوریتم‌های ماشین یادگیرنده (Ayyadevara, 2018) (Lakshmi, 2016) (درخت تصمیم، گرادیان تقویت‌شده و درخت تصمیم افزوده) (Schwartz, 2014) (Mitchell, 1997)، رویکرد و مدل جدیدی را برای تخمین تراوایی ارائه و طراحی نماییم.

در سال‌های گذشته مطالعات و پژوهش‌هایی در خصوص تخمین تراوایی مخزن با استفاده از روش‌های یادگیری ماشین انجام شده است. از جمله این پژوهش‌ها سلحشور و همکارانش در مقاله‌ای با عنوان بهینه‌سازی الگوریتم‌های یادگیری ماشین جهت تخمین تراوایی مخازن زیرزمینی نفت و گاز ارائه کردند، آن‌ها روش ماشین بردار پشتیبان با توابع هسته مختلف را جهت تخمین تراوایی استفاده کرده که در نهایت توانستند با استفاده از روش رگرسیون بردار پشتیبان (SVR)¹ و تابع هسته پایه شعاعی بهترین تخمین تراوایی را

در یکی از مخازن نفتی با ضریب همبستگی ۹۶.۱ داشته باشند (سلحشور، ۱۳۹۳). در سال ۲۰۱۶، عبدالعزیز و همکاران تحقیقی با عنوان پیش‌بینی تراوایی بهبودیافته به‌وسیله داده‌های لرزه‌نگاری و داده‌های لاگ با استفاده از تکنیک‌های هوش مصنوعی انجام دادند. در این تحقیق، یک مطالعه مقایسه‌ای در پیش‌بینی تراوایی مخزن در اثر ترکیب مجموعه داده‌های لاگ چاه و لرزه‌نگاری انجام شد. این مطالعه با به‌کارگیری سه تکنیک پیشرفته هوش مصنوعی شامل شبکه‌های فازی (FN)²، ماشین بردار پشتیبان (SVM)³ و سیستم فازی نوع دوم (T2FLS)⁴، روی مجموعه داده‌های لرزه‌نگاری و لاگ به‌صورت ترکیبی، انجام شده است. مدل‌ها برای عملکرد مطلوب بهینه‌سازی شدند و مجموعه داده‌ها به ترتیب به زیرمجموعه‌های آموزش و آزمایش به دنبال الگوی یادگیری ماشین استاندارد قرار گرفتند. نتایج حاصل از این مطالعه نشان داد که ترکیب داده‌های لرزه‌نگاری و لاگ در مقایسه با استفاده سنتی از داده‌های لاگ و لرزه-نگاری به‌صورت مجزا عملکرد بهتری داشتند. آن‌ها مدل‌های ترکیبی از سیستم فازی نوع دوم، ماشین بردار پشتیبان و شبکه‌های فازی را تحت عنوان الگوریتم انتخاب ویژگی غیرخطی ارائه دادند، مقایسه نتایج مدل‌های ترکیبی از این سه الگوریتم نشان داد که روش ترکیبی FN-SVM با ضریب تعیین ۸۲ درصد و خطای میانگین مربعات ۰.۴۶ بهترین تخمین را به دست آوردند (Fatai, 2016). طیبی در مقاله‌ای با عنوان مطالعه موردی تعیین تراوایی مخازن کربناته با استفاده از لاگ تشدید مغناطیسی هسته (NMR)⁵ در یکی از میداین جنوب غربی ایران از روش‌های نسبت تحویل رسوب (SDR)⁶ و Timur برای محاسبه تراوایی استفاده کردند دقت و صحت تخمین آن‌ها ۴۴.۱ درصد بود (طیبی، ۱۳۹۶). حکیمی نژاد و همکاران در سال ۲۰۱۷، پژوهشی بر روی داده‌های چهار چاه کنگان و دالان واقع در میدان پارس جنوبی ایران انجام دادند. آن‌ها با انتخاب ۸ ویژگی از هر چاه، با استفاده از روش انتخاب ویژگی مبتنی بر همبستگی، چهار ویژگی مؤثر در هر چاه را انتخاب کردند. سپس از روش‌های

¹ Support Vector Regression

² Fuzzy Networks

³ Support Vector Machine

⁴ Type 2 Fuzzy Logic System

⁵ Nuclear Magnetic Resonance

⁶ Sediment Delivery Ratio

دوبعدی کشف شد. این میدان بزرگ نفتی در فروفاتادگی دزفول حدود ۴۵ کیلومتری جنوب شرقی اهواز قرار دارد (میرزاقلی پور و حقی، ۱۳۶۹). روند محوری این میدان شمال غربی-جنوب شرقی است. میدان مارون در مخزن آسماری دارای طول ۶۷ کیلومتر و عرض متوسط ۵/۵ کیلومتر می باشد که بیشترین و کمترین عرض آن به ترتیب ۷ و ۳/۵ کیلومتر می باشد. موقعیت جغرافیایی این میدان نسبت به میدان های مجاور همان طور که در شکل (۱) نشان داده شده است، از شمال توسط میدان رامین، از شرق توسط میدان کوپال، از غرب و شمال غرب توسط میدان های شادگان و اهواز و از جنوب توسط میدان رامشیر محدود می شود (سراج، ۱۳۸۴).



شکل ۱. موقعیت جغرافیایی میدان نفتی مارون (Amiri Bakhtiyar et al., 2018)

۲. روش کار

در این مطالعه روش تحلیل مطابق شکل (۱) ارائه شده است. به این ترتیب که ابتدا روش های یادگیری ماشین غیر خطی که تاکنون در تخمین تراوایی کمتر استفاده شده و همچنین جزو الگوریتم های قوی در یادگیری می باشند انتخاب شده اند، سپس اطلاعات مربوط به لاگ های پتروفیزیکی در اختیار قرار داده می شود، اطلاعات مربوطه پس از استانداردسازی در بازه ۰ تا ۱، بر اساس حجم مقادیر لیتولوژی در هر رکورد، یکی از برچسب های ماسه سنگ^۸، کلسیت^۹، دولومیت^{۱۰}

رگرسیون، شبکه عصبی چندلایه، مدل درخت خطی، سیستم فازی نوع یک و سیستم فازی نوع دو برای پیش بینی تراوایی استفاده کردند. نتایج نشان داد که با توجه به وجود عدم قطعیت در پارامترهای پتروفیزیکی و تراوایی، سیستم فازی نوع دو عدم قطعیت ها را بهتر پوشش می دهد. این روش در حالت پایه، تراوایی را با دقت ۰.۹۴۸۱ و ریشه دوم میانگین مربعات خطا ۰.۳۰۶۰ پیش بینی کرد که با استفاده از روش ترکیبی الگوریتم ژنتیک و الگوریتم جستجوی گرانشی، توابع عضویت فازی بهبود یافت و این بهبودها منجر به افزایش دقت پیش بینی تراوایی با ضریب تعیین ۰.۹۷۶۸ و کاهش ریشه دوم میانگین مربعات خطا به مقدار ۰.۱۶۰۲ شد (حکیمی نژاد، ۱۳۹۷). در سال ۲۰۱۸، احمدی و همکاران، مقایسه روش های یادگیری ماشین برای تخمین تراوایی و تخلخل مخازن نفتی با استفاده از نگاره های پتروفیزیکی، روش ترکیبی *HGAPSO-LSSVM* را ارائه دادند (Ahmadi & Zhangxing, 2018)، در سال ۲۰۱۹، ادنیران یک مدل گروهی رقابتی برای پیش بینی تراوایی در مخازن ناهمگن نفت و گاز ارائه کرد نتایج نشان داد که مدل های گروهی رقابتی، دقت پیش بینی بهتری در تخمین تراوایی دارد، آن ها از یک مدل گروهی برای تخمین تراوایی شامل ماشین های شبکه عصبی مصنوعی (*ANN*) و ماشین بردار پشتیبان (*SVM*) و سیستم های تطبیقی استنتاج عصبی فازی (*ANFIS*) بهره بردند که موفق شدند با ضریب همبستگی ۹۵.۳۴ درصد تراوایی را تخمین بزنند (Adeniran, 2019)، سباستین و همکاران در مقاله تخمین تراوایی مطلق با استفاده از شبکه های عصبی بر اساس لاگ های چاه و داده های آزمایشگاهی در سال ۲۰۱۹، توانستند با ضریب همبستگی ۷۳ درصدی بین تراوایی حاصل از داده های آزمایشگاه و تراوایی حاصل از شبکه عصبی کار خود را ارائه دهند (Waszkiewicz, 2019).

۱.۱. موقعیت جغرافیایی و ویژگی های ساختمانی

میدان نفتی مارون

میدان نفتی مارون در سال ۱۹۶۳ به روش لرزه نگاری

⁷ Adaptive network-based fuzzy inference system

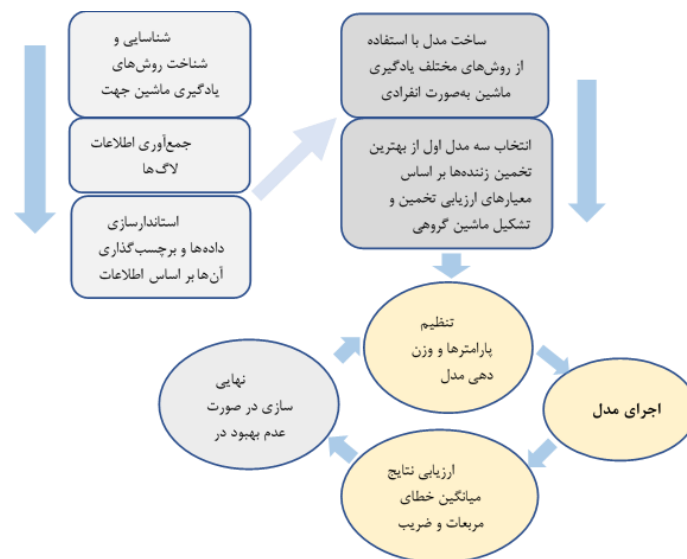
⁸ Sand Stone

⁹ Calcite

¹⁰ Dolomite

ماشین گروهی انتخاب می‌شوند که معیار ارزیابی عملکرد کارایی مدل‌ها، ضریب همبستگی بین مقادیر تخمین زده‌شده و مقادیر واقعی تراوایی و همچنین میانگین مربعات می‌باشد. از اینجا به بعد مدل ترکیبی گروهی ساخته‌شده به ازای مقادیر مختلف پارامترهای ورودی و همچنین وزن دهی متفاوت به مدل‌های درون ماشین که سه مدل می‌باشد ارزیابی خواهد شد و بهترین عملکرد ماشین به‌عنوان مدل نهایی طراحی شده برگزیده خواهد شد.

شیل^{۱۱} و انیدریت^{۱۲} را به خود اختصاص می‌دهد و داده‌ها بر این اساس تفکیک می‌شوند. آنگاه به ازای هر خوشه، مدل-های یادگیری ماشین کی-نزدیکترین همسایه (KNN)^{۱۳}، رگرسیون جنگل تصادفی (RFR)^{۱۴}، رگرسیون درخت تصمیم (DTR)^{۱۵}، رگرسیون گاوسی تقویت‌شده (GBR)^{۱۶}، رگرسیون سازگار تقویت‌شده (ABR)^{۱۷} و رگرسیون درخت افزوده (ETR)^{۱۸} با پارامترهای مختلف در هر مدل و در تکرارهای مشابه بررسی خواهد شد. سپس سه مدل از الگوریتم‌هایی که بهترین عملکرد را دارند برای تشکیل یک



شکل ۱. گام‌های تحلیل مسئله

آب^{۲۲}، فتوالکتریک^{۲۴}، تخلخل مؤثر^{۲۵} و اطلاعات لیتولوژی شامل حجم کلسیت، حجم دولومیت، حجم ماسه‌سنگ، حجم شیل، حجم انیدریت، از لاگ‌های چاه‌های به همراه

۱.۲ شناخت و آماده‌سازی داده‌ها

در این مطالعه ابتدا تعداد ۷۷۸۵ رکورد مربوط به ویژگی‌های پرتو گاما^{۱۹}، چگالی^{۲۰}، مقاومت الکتریکی^{۲۱}، نوترون^{۲۲}، اشباع

11 Shale
 12 Anhydrite
 13 K Nearest Neighbor Regressor
 14 Random Forest Regressor
 15 Decision Tree Regressor
 16 Gaussian Boost Regressor
 17 Adaptive Boosting Regressor
 18 Extra Tree Regressor
 19 Gamma Ray
 20 Density
 21 Resistivity Environmentally
 22 Neutron
 23 Water Saturation
 24 Photoelectric Factor
 25 Effective Porosity

شباهت بین اشیاء، باید برچسب‌هایی به‌طور خودکار نسبت داده شوند. درحالی‌که در دسته‌بندی برچسب‌های اولیه موجود است و باید با استفاده از الگوهای پیش‌بینی قادر به برچسب‌گذاری برای مشاهدات جدید باشیم (Aggarwal & Reddy, 2014).

اگرچه بیشتر الگوریتم‌ها یا روش‌های خوشه‌بندی مبنای یکسانی دارند ولی تفاوت‌هایی در شیوه اندازه‌گیری شباهت یا فاصله و همچنین انتخاب برچسب برای اشیاء هر خوشه در این روش‌ها وجود دارد. به‌طور کلی چهار گروه اصلی برای الگوریتم‌های خوشه‌بندی وجود دارد. الگوریتم‌های خوشه‌بندی تفکیکی، الگوریتم‌های خوشه‌بندی سلسله مراتبی، الگوریتم‌های خوشه‌بندی بر مبنای چگالی و الگوریتم‌های خوشه‌بندی بر مبنای مدل (Lakmi, 2016). در این مطالعه روش خوشه‌بندی کی-میانگین (خوشه‌بندی تفکیکی)، مورد بررسی قرار گرفته است (Wang, 2003).

۳.۲ برچسب‌گذاری داده‌ها برحسب اطلاعات

لیتولوژی

در این قسمت داده‌های ورودی را به ازای هر رکورد مطابق جدول (۱) برچسب‌گذاری می‌کنیم. بدین گونه که اطلاعات لیتولوژی شامل ستون‌های حجم کربنات، حجم دولومیت، حجم ماسه‌سنگ، حجم شل در هر نمونه را بررسی کرده و چنانچه مقدار هر کدام از آن‌ها از سایر ستون‌های دیگر بیشتر باشد برچسب آن رکورد را مطابق ستون گونه سنگ در جدول (۱) به خود اختصاص می‌دهد. درنهایت داده‌ها در چهار خوشه تقسیم‌بندی می‌شوند که هر خوشه به‌صورت مجزا برای مدل تخمین زنده به‌عنوان ورودی انتخاب خواهد شد.

جدول ۱ شمای کلی از برچسب‌های اطلاعات ورودی

عمق لاگ	عمق مغزه	حجم کلسیت	حجم دولومیت	حجم ماسه	حجم شیل	حجم انیدریت	گونه سنگ
۲۹۵۸.۵۴۱	۲۹۵۸.۲۵	۰.۸۵۳	۰.۱۴۷	۰	۰	۰	Calcite
۲۹۶۰.۶۷۵	۲۹۶۰.۷۱	۰.۴۷۴۶	۰.۵۲۵۴	۰	۰	۰	Dolomite
۳۰۲۲.۹۲۱	۳۰۲۲.۵۳	۰.۲۳۳۶	۰	۰.۷۶۶۴	۰	۰	Sand
۳۱۱۳.۰۷۵	۳۱۱۳.۱۱	۰	۰.۰۱۶۳	۰	۰.۹۸۳۷	۰	Shale

تراوایی که از داده‌های مغزه پس از هم عمق سازی با داده‌های لاگ از همان چاه به‌دست آمده، مشتمل بر تعداد ۱۴ چاه از مخزن نفتی مارون ایران انتخاب شد.

مجموعه‌ای آموزشی به‌صورت تصادفی از روی داده‌های لاگ-های مربوط به ۱۴ چاه انتخاب شده و مجموعه آزمایشی نیز به‌صورت تصادفی از بین داده‌های کل ۱۴ چاه انتخاب شده است. در هر نمونه آزمایشی و آموزشی دو مجموعه داده وجود دارد، که درنهایت به‌صورت زیر دسته‌بندی خواهند شد:

$$\{(x_1, y_2), \dots, (x_1, y_2) \in R^N, y \in \} \quad (1)$$

که به‌صورت زیر می‌باشند:

$$Xi = (DT, CGR, PEF, RHOB, NPHI, RT, SW, POR) \text{ و } Yi = Perm$$

بدین ترتیب که تراوایی به‌عنوان پارامتر خروجی و به مقادیر لاگ پرتو گاما، تخلخل، پرتو الکتریکی، چگالی، نوترون، سونیک، مقدار اشباع، فتوالکتریک به‌عنوان بردار ورودی می‌باشد.

۲.۲ خوشه‌بندی

تحلیل خوشه‌ای^{۲۶} یا به‌طور خلاصه خوشه‌بندی، فرآیندی است که به کمک آن می‌توان مجموعه‌ای از اشیاء را به گروه‌های مجزا افراز کرد. هر افراز یک خوشه نامیده می‌شود. اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه هستند و در عوض میزان شباهت بین خوشه‌ها کمترین مقدار است. در چنین حالتی هدف از خوشه‌بندی، نسبت دادن برچسب‌هایی به اشیاء است که نشان‌دهنده عضویت هر شیء به خوشه است. در تحلیل خوشه‌ای بر خلاف دسته‌بندی^{۲۷}، به دلیل عدم وجود برچسب‌های اولیه به‌عنوان یک روش غیر نظارتی می‌باشد. در این روش، بر اساس ویژگی‌های مشترک و روش‌های اندازه‌گیری فاصله یا

²⁶ Cluster Analysis

²⁷ Classification

۴.۲ انتخاب روش استانداردسازی داده‌ها

روش‌های مختلفی برای استانداردسازی وجود دارد که در تمام این روش‌ها داده‌ها طوری تبدیل می‌شوند که شرایط مورد نظر تأمین گردد. یکی از مهم‌ترین روش‌های استانداردسازی، تبدیل داده‌ها به مجموعه جدیدی است که در آن، همه‌ی مقادیر بین صفر و یک می‌باشند:

$$Z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

که در آن z_i مقدار استانداردشده، x_i مقدار هر یک از داده‌ها، x_{min} حداقل کل داده‌ها x_{max} حداکثر مقدار داده‌ها می‌باشد.

۵.۲ روش ارزیابی استفاده‌شده

نتایج به‌دست‌آمده از الگوریتم‌های مورد استفاده از لحاظ دقت، صحت، مقدار و کیفیت مورد بررسی قرار گیرند تا بتوان بهترین حالت برای مدل‌سازی را انتخاب نمود لذا روش به دست آوردن این کمیت‌ها در زیر توضیح داده‌شده است.

خطای میانگین مربعات^{۲۸}:

(۳)

$$MSE = (\sum Actual - Forecast)^2 / n = \sum (Error)^2 / n$$

رابطه‌ی ۳-، خطای میانگین مربعات است که در آن n ، تعداد نمونه‌های تخمینی، $Actual$ ، مقدار واقعی هر برچسب و $Forecast$ ، مقدار تخمینی برای هر یک است. هر چه مقدار کمیت MSE کمتر باشد، ضریب خطا نیز کمتر خواهد بود. ضریب همبستگی، شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) را نشان می‌دهد. این ضریب بین (۱) تا (-۱) است و در حالت عدم وجود رابطه بین دو متغیر، برابر صفر می‌باشد. همبستگی بین دو متغیر تصادفی X و Y به‌صورت رابطه‌ی ۴- تعریف می‌شود:

(۴)

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

که در آن E عملگر امید ریاضی، cov به معنای کوواریانس، $corr$ نماد معمول برای همبستگی پیرسون و σ نماد انحراف معیار است.

۶.۲ آموزش و ساخت مدل

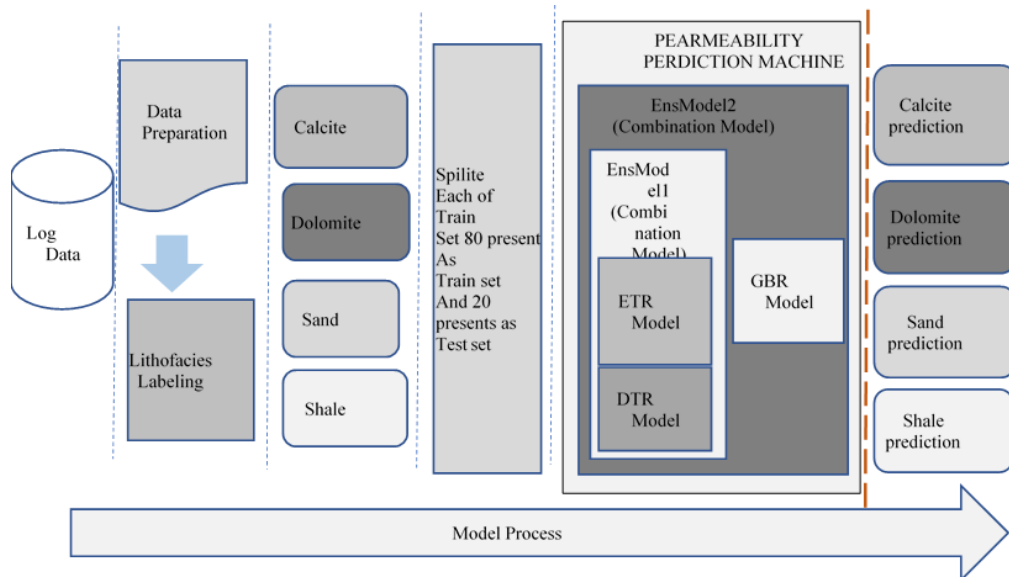
در این مرحله ابتدا داده‌ها را بر اساس اطلاعات برچسب‌گذاری شده در بخش ۲-۳ به چهار خوشه A, B, C, D تقسیم‌بندی می‌کنیم و هر کلاس را به دو مجموعه ۸۰ درصدی آموزش و ۲۰ درصدی آزمایش به‌صورت تصادفی تقسیم کرده، سپس با استفاده فریم‌ورک و زبان برنامه‌نویسی پایتون مدل‌های یادگیری ماشین رگرسیونی درخت تصمیم، درخت تصمیم افزوده، رگرسیون جنگل تصادفی، کی-نزدیکترین همسایه، رگرسیون انطباقی تقویت‌شده، رگرسیون گرادیانی تقویت‌شده، مدل‌های داده را به‌صورت انفرادی آموزش دادیم که نتایج عملکردی مطابق جدول (۲) به ترتیب برای هر خوشه به دست آمد.

با توجه به نتایج جدول (۲) مدل‌های ETR و GBR و DTR برای طراحی مدل گروهی تخمین زنده تراوایی انتخاب شد. سپس مدل ترکیبی با استفاده از ترکیبات مختلف و وزن‌های متفاوت از مدل‌های ETR و GBR و DTR در تکرارهای مختلف ایجاد گردید، نتایج حاصل از اجراهای مختلف توسط شاخص‌های ارزیابی R^2 و MSE در هر تکرار بررسی شد و با توجه به نتایج بهترین تکرار، مدلی ترکیبی به‌صورت ترکیب وزنی ۹۰ درصد از روش ETR و ۱۰ درصدی روش DTR به‌عنوان مدل گروهی اول تشکیل شده که با نام $Ensemble1$ نام‌گذاری شد و در مرحله بعد نیز با روش GBR با درصد وزنی ۱۰ به ۹۰ ترکیب شده و با نام $Ensemble2$ که در شکل (۲) به نمای کلی مدل طراحی‌شده، اشاره می‌شود. درصد وزنی ترکیب مدل‌ها توسط حلقه تکرار و بررسی همه حالات مطابق با فرایند شکل (۱) به‌دست‌آمده است که ارزیابی خروجی حالات مختلف توسط محاسبه ضریب همبستگی و خطای میانگین مربعات در هر تکرار محاسبه‌شده و در نهایت بهینه‌ترین برآزش به‌عنوان مدل نهایی مطابق شکل (۲) برگزیده شد.

²⁸ MEAN SQUARE ERROR

جدول ۲ نتایج اجرای روش های یادگیری شش گانه بر روی ۴ خوشه ورودی

خوشه <i>D</i>	خوشه <i>C</i>	خوشه <i>B</i>	خوشه <i>A</i>	معیار ارزیابی	الگوریتم یادگیری ماشین
تعداد: ۲۸۵۰	تعداد: ۹۵۶	تعداد: ۳۵۶۷	تعداد: ۴۱۲		
۸۸.۰۵	۷۴.۵	۸۱.۷	۸۳.۶	R^2	ETR
$31(e-0.6)$	$42(e-0.6)$	$36(e-0.6)$	$34(e-0.6)$	MSE	
۷۱.۸۸	۵۳.۳۷	۶۰.۱۳	۷۹.۵	R^2	RFR
$68(e-0.6)$	$86(e-0.5)$	$54(e-0.5)$	$63(e-0.6)$	MSE	
۷۳.۶	۶۴.۹	۵۵.۷	۷۵.۴	R^2	KNR
$69(e-0.6)$	$51(e-0.6)$	$81(e-0.5)$	$69(e-0.6)$	MSE	
۴۷.۰۲	۲۳.۸۹	۴۹.۴۴	۳۸.۶۶	R^2	ABR
$79(e-0.5)$	$94(e-0.4)$	$81(e-0.5)$	$89(e-0.5)$	MSE	
۷۸.۱۱	۷۸.۲۲	۶۸.۳۳	۶۴.۰	R^2	GBR
$39(e-0.6)$	$38(e-0.6)$	$55(e-0.5)$	$51(e-0.6)$	MSE	
۸۲.۹	۷۱.۳	۷۷.۲۲	۹۶.۲	R^2	DTR
$38(e-0.6)$	$68(e-0.6)$	$41(e-0.6)$	$8.1(e-0.6)$	MSE	



شکل ۲. شمای اجرا شده در آموزش مدل و تخمین تراوایی

بدون اینکه تفکیکی انجام شود توسط مدل گروهی آموزش داده شد به نحوی که ۸۰ درصد داده ها به عنوان مجموعه آموزش و ۲۰ درصد داده ها به عنوان مجموعه آزمایشی

در مرحله بعد، جهت مقایسه نتایج در حالت پیش از تفکیک گونه های سنگی، داده های اولیه شامل تعداد ۷۷۸۵ نمونه که بر اساس مقدار خصوصیت عمق نمونه مرتب سازی شده و

مدل ترکیبی گروهی شامل سه روش ETR، DTR و GBR مطابق شکل (۲)، طراحی شد که در جدول (۴) نتایج اجرای مدل به تفکیک برچسب‌گذاری داده‌ها و میزان کمیت هر خوشه، همچنین ضریب تعیین و خطای میانگین مربعات و خطای مطلق میانگین آورده شده است. در شکل (۳) میزان تراوایی تخمین زده شده بر اساس تراوایی واقعی، پیش از خوشه‌بندی و تفکیک گونه‌های سنگی بر اساس اطلاعات لیتولوژی نمایش داده شده است که نشان‌دهنده تأثیر تفکیک گونه‌های سنگی در دقت تخمین می‌باشد

انتخاب شدند و پس از اجرای مدل گروهی، نتایج در جدول (۳) نمایش داده شده است.

جدول ۳. نتایج اجرای مدل گروهی تخمین تراوایی قبل از خوشه‌بندی

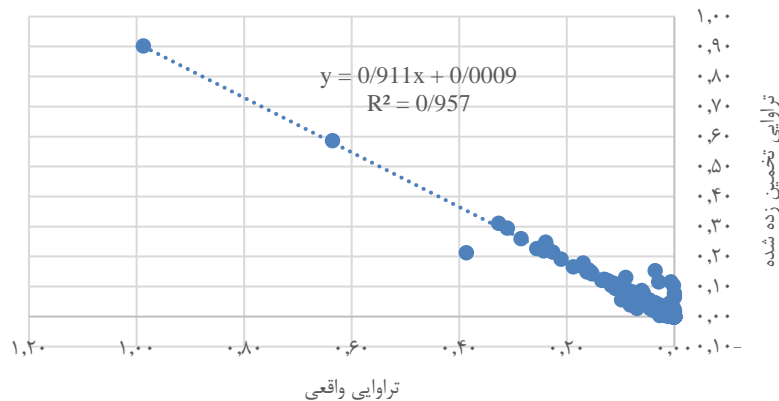
تعداد داده‌های مجموعه آزمایش	R^2	MSE	MAE
۱۵۵۶	۹۵.۷	$۷.۶۱۰۱۵(e-۵)$	۰.۰۰۱۶۶۸

۳. نتایج

در این مقاله با استفاده از روش‌های یادگیری شش‌گانه یک

جدول ۴ نتایج اجرای مدل گروهی تخمین تراوایی

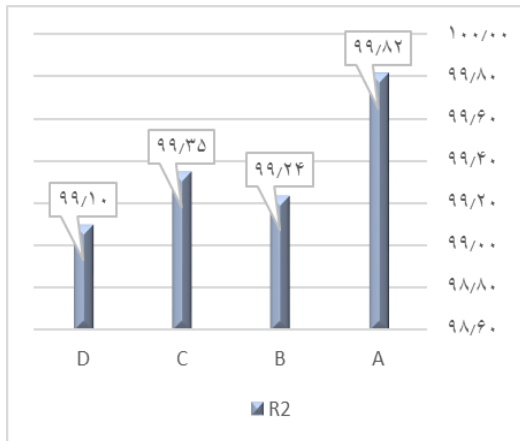
نام خوشه	تعداد نمونه در خوشه	R^2	MSE	MAE
A	۴۱۲	۹۹.۸۲	$۲.۲۵۰۸۴(e-۶)$	۰.۰۰۰۵۶۷۹۳۱
B	۳۵۶۷	۹۹.۲۴	$۳.۰۶۷۵(e-۶)$	۰.۰۰۰۵۰۰۹۷۲
C	۹۵۶	۹۹.۳۵	۰.۰۰۰۱۹۲۰	۰.۰۰۰۵۰۶
D	۲۸۵۰	۹۹.۱	$۲.۱۰۹۲(e-۶)$	۰.۰۰۰۵۰۳۴۱



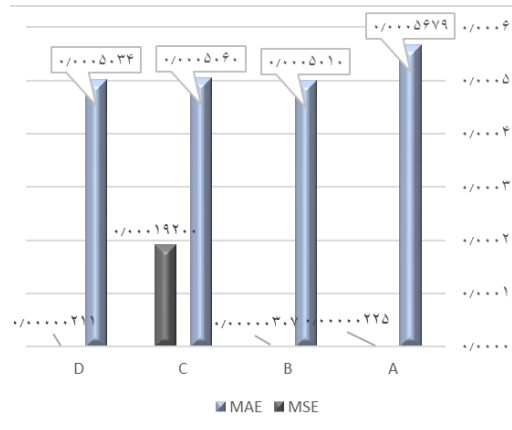
شکل ۳ نمودار تخمین تراوایی با استفاده از مدل گروهی پیش از تفکیک گونه‌های سنگی

تراوایی تخمین زده شده در مقابل میزان واقعی تراوایی برای خوشه B نمایش داده شده است. در شکل (۸) تراوایی تخمین زده شده در مقابل میزان واقعی تراوایی برای خوشه C نمایش داده شده است. در شکل (۹) تراوایی تخمین زده شده در مقابل میزان واقعی تراوایی برای خوشه D نمایش داده شده است.

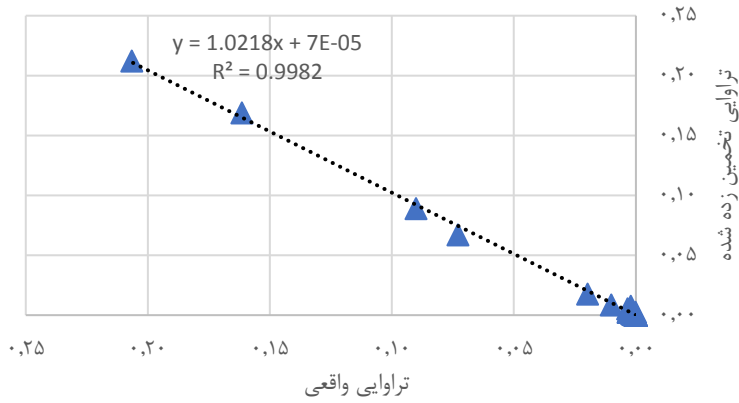
در شکل (۴) خطای میانگین مربعات و خطای مطلق مربعات بر اساس خوشه‌ها نمایش داده شده است. در شکل (۵) ضریب همبستگی جهت تخمین تراوایی بر اساس هر خوشه نمایش داده شده است. در شکل (۵) تراوایی تخمین زده شده در مقابل میزان واقعی تراوایی برای خوشه A نمایش داده شده است. در شکل (۶)



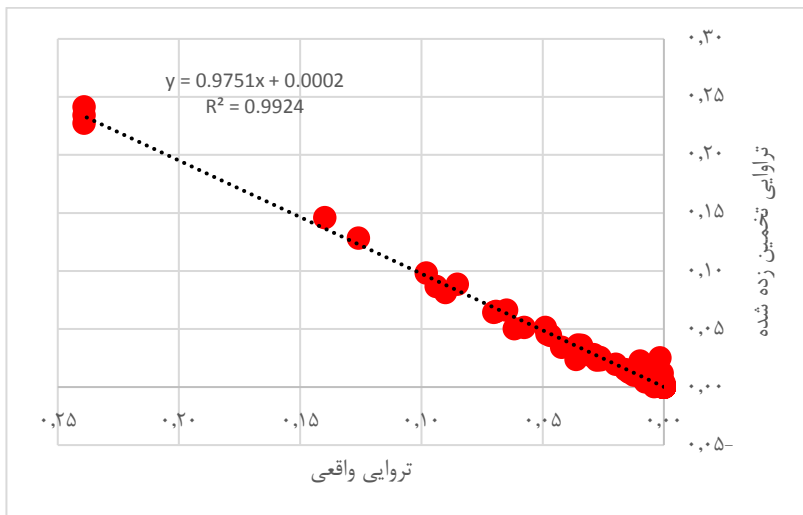
شکل ۵. نمودار مقایسه R² برای هر خوشه



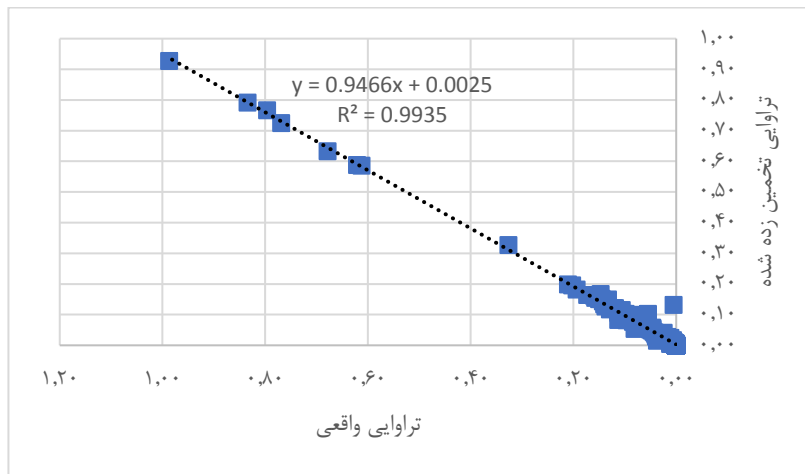
شکل ۴. نمودار مقایسه خطا برای هر خوشه



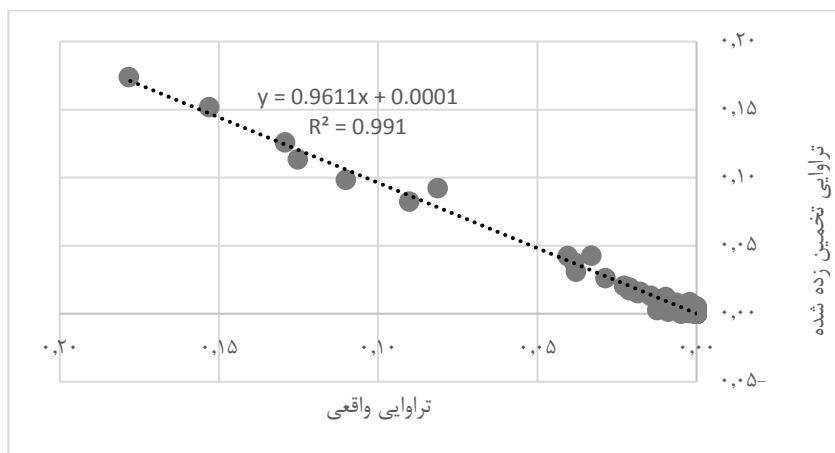
شکل ۶. نمودار تخمین تراوایی برای خوشه A (شل)



شکل ۷. نمودار تخمین تراوایی برای خوشه B (کلسیت)



شکل ۸. نمودار تخمین تراوایی برای خوشه C (ماسه سنگ)



شکل ۹. نمودار تخمین تراوایی برای خوشه D (دولومیت)

۴. بحث و بررسی

بررسی نتایج جدول شماره (۴) نشان می‌دهد که مدل ترکیبی گروهی نتایج بسیار خوبی برای همه خوشه‌ها که شامل تخمین تراوایی با ضریب تعیین میانگین ۹۹.۳۷۷ و میانگین خطای میانگین مربعات ۰.۰۰۰۰۴۹۸ بوده و بهترین تخمین را برای خوشه A با شاخص ضریب تعیین ۹۹.۸۲ و خطای میانگین مربعات (e-۰۶) ۲.۲۵۰۸۴ حاصل شده است. همچنین با نگاهی بر نتایج دسته‌های دیگر یک عامل جهت اختلاف بین دقت تخمین، تعداد نمونه‌های هر خوشه می‌باشد، البته کیفیت مقادیر نمونه‌ها نیز در تخمین بی‌تأثیر نبوده است. همچنین جهت برازش مدل با استفاده از شاخص ریشه میانگین

مربعات باقیمانده ($RMSE^2$)، چنانچه این شاخص کمتر از ۰.۰۵ باشد نشان‌دهنده نیکویی برازش مدل بر داده‌ها است و مطابق جدول (۵)، این موضوع به‌وضوح مشخص است. از سوی دیگر همان‌طور که در نمودارهای ۶ تا ۹ نمایش داده شده، داده‌ها در هر خوشه متنوع بوده و بیشتر تمرکز داده از لحاظ کمیت بین ۰ تا ۰.۰۵ می‌باشد. نتایج دقت تخمین این پژوهش در مقایسه با مطالعات بخش ۱ بیانگر توانایی بالای مدل طراحی شده نسبت به سایر روش‌ها است.

²⁹ Root Mean Square Residual

جدول ۵ برازش مدل بر خوشه‌ها

نام خوشه	تعداد نمونه‌های خوشه	ریشه میانگین مربعات باقیمانده
<i>A</i>	۴۱۲	۰.۰۰۱۵
<i>B</i>	۳۵۶۷	۰.۰۰۱۷
<i>C</i>	۹۵۶	۰.۰۱۳۸
<i>D</i>	۲۸۵۰	۰.۰۰۱۴

۵. نتیجه‌گیری

بسیار وابسته به اطلاعات و مجموعه داده ورودی مدل است و برای مخازن مختلف، می‌بایست پیکربندی و تنظیمات پارامترهای مدل به صورت خاص انجام گیرد.

در این پژوهش برای پیش‌بینی تراوایی از روش‌های یادگیری ماشین به عنوان راهکاری جایگزین برای روش‌های تجربی و آزمایشگاهی استفاده شد و ضمن طراحی یک مدل گروهی جهت تخمین تراوایی، نتایج بسیار خوب زیر نیز به دست آمد: تخمین تراوایی با دقت بسیار بالا و با ضریب همبستگی ۹۹.۷۶ درصد و ضریب خطا به میزان $(e-06) 2.25 \times 10^{-6}$ استفاده از روش گروهی مدل‌های یادگیری ماشین که باعث تقویت نتایج تخمین خواهد شد. تاثیر تفکیک سنگ‌ها بر اساس اطلاعات لیتولوژی و ارتباط مؤثر آن با نتایج طراحی مدل‌های یادگیری جهت تخمین تراوایی

۶. سپاس‌گزاری

نویسندگان بر خود لازم می‌دانند از مدیران و مسئولین شرکت ملی مناطق نفت‌خیز جنوب به خصوص اداره مهندسی پتروفیزیک به سبب همکاری و در اختیار گذاشتن داده‌های مورد استفاده تشکر نمایند.

۷. مراجع

- North, F.K (1985) Petroleum Geology, Allen & Unwin.
- Ramzi, H.R (1998) Well Logging, SANAM Publishing.
- Oberto, S. (1984) Fundamentals of Well-Log Interpretation – The Acquisition of Logging Data, Elsevier.
- Ayyadevara, V. (2018) Pro Machine Learning Algorithms, Apress.
- Abbaszadeh, M., Fuji, H. and Fujimoto, F.,)1996(Permeability prediction by hydraulic flow units-theory and applications. SPE Formation Evaluation, Vol. 11, No. 4, pp 263- 271.
- Lakhmi, C. Jain (2016) Foundations and Methods in Combinatorial and Statistical Data Analysis and Clustering, Springer Press.
- Shalev-Schwartz, Sh. (2014) understanding-machine-learning-theory-algorithms, Cambridge University Press.
- Aggarwal, C. C. & Reddy, C. K. (eds.) (2014). Data Clustering: Algorithms and Applications. CRC Press.

- Wang, J. (2003). *Data Mining: Opportunities and Challenges*, USA Montclair, IRM Press.
- Mitchell, M. (1997) *Machine-Learning-Tom-Mitchell*, McGraw-Hill.
- Ahmadi, Mohammad. ali. & Zhangxing, Chen (2018), Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via Petro-physical logs, *KeAi Elsevier*, 6, 2:831-844.
- Adeniran A. (2019) A competitive ensemble model for permeability prediction in heterogeneous oil and gas reservoirs, *Elsevier Paper*, 4, 80:353-367.
- Waszkiewicz, S., rakowska-Madejska, P., Puskarczyk, E. (2019), Estimation of absolute permeability using artificial neural networks (Multilayer perceptron) based on well logs and laboratory data from Silurian and Ordovician deposits in SE Poland, *Springer*, 6, 67:1885–1894.
- Fatai, A., Abdulazeez, A., Abdullatif, A. (2016), Improved Permeability Prediction from Seismic and Log Data using Artificial Intelligence Techniques, *SPE 164465*, pp 1-3.
- Amiri Bakhtiyar, H., A. Telmadarreie, M. Shayesteh, M. H. Heidari Fard, H. Talebi, and Z. Shirband. (2018). Estimating Total Organic Carbon Content and Source Rock Evaluation, Applying $\Delta\log R$ and Neural Network Methods: Ahwaz and Marun Oilfields SW of Iran. *Petroleum Science and Technology*, Vol. 29, No. 16, pp 1691–1704. <https://doi.org/10.1080/10916461003620495>
- ایزیدی م. (۱۳۹۱) "محاسبه تراوایی در چاه‌های فاقد مغزه با استفاده از مفهوم واحدهای جریان‌ی هیدرولیکی"، مجله اکتشاف و تولید نفت، ۱۱، ۹۸: صفحات ۶۸-۶۵.
- سلحشور ع.، سهیلی ف.، کمری م. (۱۳۹۳) "بهینه‌سازی الگوریتم‌های یادگیری ماشین جهت تخمین تراوایی مخازن زیرزمینی نفت و گاز"، مجله اکتشاف و تولید نفت، ۱۲، ۳: صفحات ۶۲-۵۶.
- طیبه ه.، حبیب نیا ب. (۱۳۹۶) "مطالعه موردی تعیین تراوایی مخازن کربناته با استفاده از لاگ NMR در یکی از میادین جنوب غربی ایران"، مجله علمی پژوهشی زمین‌شناسی نفت ایران، ۷، ۱۳: صفحات ۶۲-۴۳.
- حکیمی نژاد ح.، میرزازایی م.، نجار اعرابی ب. (۱۳۹۷) "پیش‌بینی تراوایی سنگ مخزن با استفاده از روش‌های عدم قطعیت سیستم فازی نوع دو"، مجله علمی پژوهشی ژئوفیزیک ایران، ۱، ۱۲: صفحات ۹۱-۸۲.
- میرزا قلی پور، علی؛ حقی، عبدالحمید (۱۳۶۹). مطالعه زمین‌شناسی میدان نفتی مارون، گزارش شماره پ-۴۲۱۰، اداره کل زمین‌شناسی گسترشی، شرکت ملی مناطق نفت‌خیز جنوب، ص ۵۵.
- سراج، م. (۱۳۸۴). تحلیل ساختاری مقدماتی میادین نفتی مناطق نفت‌خیز جنوب (محدوده فروافتادگی دزفول شمالی)، گزارش شماره پ-۵۶۱۳، شرکت ملی مناطق نفت‌خیز جنوب.