

# تخمین لاگ صوتی برشی با استفاده از روش‌های یادگیری ماشین، و مقایسه با داده‌های حاصل از مغزه با استفاده از نرم‌افزار پایتون

هوشنگ مهرابی<sup>(۱\*)</sup>، ابراهیم سفیداری<sup>(۲)</sup>، سیده سپیده میرربیع<sup>(۳)</sup>، صادق براتی بلداجی<sup>(۱)</sup> و سید محمد زمانزاده<sup>(۴)</sup>

۱. کارشناس ارشد، گروه زمین‌شناسی نفت، پژوهشکده علوم پایه کاربردی، جهاد دانشگاهی، تهران، ایران
۲. استادیار گروه زمین‌شناسی نفت، پژوهشکده علوم پایه کاربردی، جهاد دانشگاهی، تهران، ایران
۳. دانش‌آموخته کارشناسی ارشد، دانشکده زمین‌شناسی، پردیس علوم، دانشگاه تهران، تهران، ایران
۴. دانشیار گروه سافت راک، دانشکده زمین‌شناسی، پردیس علوم، دانشگاه تهران، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۸/۰۶

تاریخ پذیرش: ۱۴۰۲/۱۱/۱۰

## چکیده

امروزه استفاده از روش‌های یادگیری ماشین در تخمین داده‌های پتروفیزیکی کاربرد فراوانی دارد. در این مطالعه سعی شده لاگ صوتی برشی از دیگر داده‌های پتروفیزیکی با روش‌های یادگیری ماشین محاسبه و با داده‌های صوتی حاصل از مغزه مقایسه شوند. برای این منظور در ابتدا برای نرمال‌سازی داده‌ها روش‌های محاسباتی انحراف معیار، جنگل انزوا، حداقل کوواریانس و فاکتور پرتی استفاده و مورد مقایسه قرار گرفتند. با توجه به مقدار داده‌های حذفی و نمودارهای جعبه‌ای رسم شده، روش انحراف معیار برای نرمال‌سازی انتخاب شد. روش‌های یادگیری ماشین استفاده شده شامل جنگل تصادفی، رگرسیون چندگانه، رگرسیون تقویت شده، رگرسیون بردار پشتیبان، K نزدیک‌ترین همسایه و پرسپترون چندلایه می‌باشند. رگرسیون چندگانه پایین‌ترین اندیس ارزیابی ( $R^2 = 0/94$ ) را داشت در حالی که رگرسیون جنگل تصمیم‌گیری با اندیس ارزیابی برابر  $0/98$ ، بالاترین همبستگی بین لاگ‌های صوتی برشی تخمین زده شده و لاگ صوتی اصلی را داشتند. بنابراین برای تخمین نهایی از رگرسیون جنگل تصادفی استفاده شد و برای جلوگیری از کلیت بخشی یا بیش‌برازش داده‌ها، از تابع GridSearchCV هایپارامترهای بهینه محاسبه و تخمین نهایی انجام شد. لاگ صوتی برشی تخمینی، با داده‌های صوتی حاصل از مغزه قرابت بسیار بالایی را نشان دادند.

واژه‌های کلیدی: پایتون، تخمین، لاگ صوتی برشی، یادگیری ماشین.

## مقدمه

محاسبات تزریق گاز  $CO_2$  و عملیات شکافت هیدرولیکی کاربرد دارند. برای محاسبه پارامترهای ژئومکانیکی مانند ویژگی‌های الاستیک سنگ، استرس درجا، فشار منفذی از داده‌های لاگ موجود استفاده می‌کنند؛ یکی از مهم‌ترین

پارامترهای مکانیک‌سنگی مخزن از مهم‌ترین داده‌های مورد نیاز در میدان‌های نفت و گاز می‌باشند. این داده‌ها در بخش‌های مختلف مطالعه مخزن از اکتشاف تا تولید،

\* نویسنده مرتبط: Houshangmehrabi@gmail.com

به حداقل رساندن مجموع اختلاف مجذور فاصله بین نقطه‌های داده واقعی و موارد پیش‌بینی شده پیدا می‌کند (Goldberger, 1962; Forkuor et al., 2017). بسته به تعداد ورودی‌ها، این تکنیک را می‌توان رگرسیون خطی ساده (فقط یک ورودی) یا چندگانه (دارای بیش از یک ورودی) نامید. شبکه عصبی مصنوعی (ANN) به‌عنوان یک ابزار مدل‌سازی داده‌های آماری غیرخطی سعی در شبیه‌سازی رفتار یک سیستم متشکل از نورون‌ها دارد و می‌تواند روابط پیچیده بین ورودی‌ها و خروجی‌ها را مدل‌سازی کند یا الگوهایی را بین آنها بیابد (Akhundi et al., 2014). شبکه‌های عصبی ممکن است دارای لایه‌های ورودی، لایه‌های پنهان و لایه‌های خروجی باشند (Heiat, 2002). در این مقاله برای تخمین لاگ صوتی برشی از روش‌های هوشمند یادگیری ماشین مانند support vector regression، random forest، XGBoost و Linear regression در محیط برنامه‌نویسی پایتون استفاده شده است. برای ارزیابی و انتخاب بهترین مدل از پارامترهای  $R_2$  و RMSE که دقت لاگ تخمین زده شده را نشان خواهند داد، استفاده شده است. مدل توضیح داده شده در بخش مخزنی سازندهای کنگان و دالان در یکی از میدان‌های دریایی دور از ساحل خلیج فارس انجام شده است. نتیجه‌های حاصل از این مطالعه می‌تواند در برآورد پارامترهای مکانیک‌سنگی و مدل‌سازی ژئومکانیکی با دقت بالا مورد استفاده قرار گیرد.

## روش مطالعه

در این مطالعه داده‌های پتروفیزیکی از دوچاه در یکی از میدان‌های نفتی خلیج فارس مورد استفاده قرار گرفته است. چاه شماره A حاوی نگارهای پتروفیزیکی از فلهلیان تا کنگان-دالان می‌باشد. با این وجود، نگار صوتی برشی فقط در سازندهای فلهلیان، هیث، سورمه، نیریز و دشتک موجود است و در سازندهای کنگان و دالان نگار صوتی برشی اندازه‌گیری نشده است. در چاه شماره B، از سازندهای کنگان و دالان نگارهای پتروفیزیکی مرسوم موجود است. همچنین در این چاه (B) اطلاعات اندازه‌گیری آزمایشگاهی

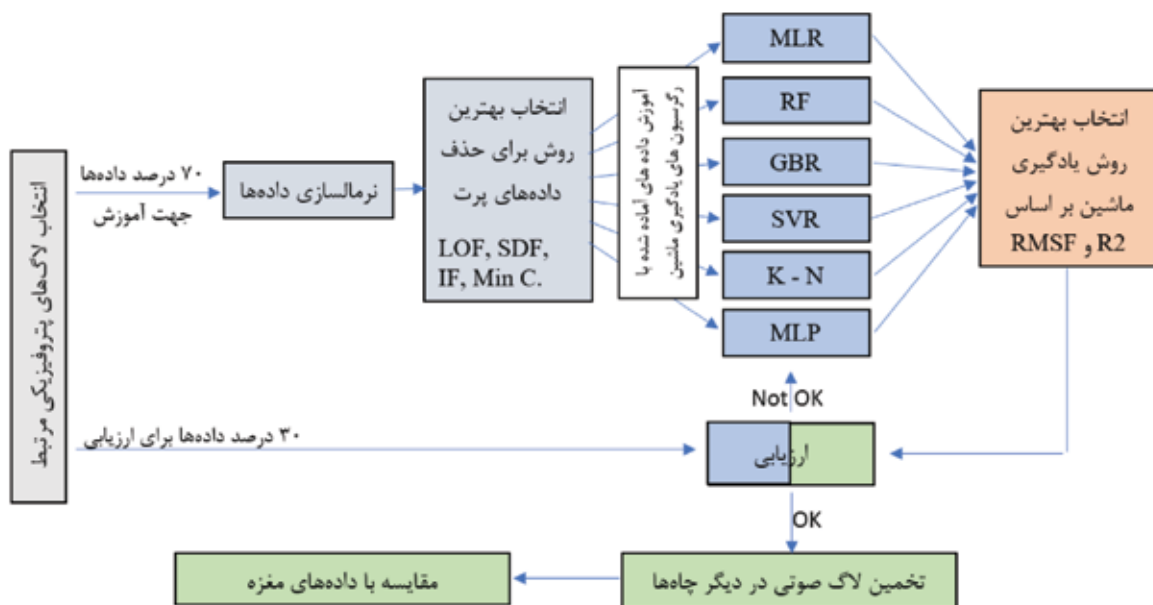
آنها داده‌های سرعت برشی (Vs) می‌باشد. با توجه به هزینه بالای تهیه لاگ‌های صوتی به‌طور معمول در همه چاه‌ها وجود ندارد. همچنین در بیشتر چاه‌های قدیمی نیز تهیه نشده‌اند. از آنجایی که لاگ‌های مرسوم پتروفیزیکی در اصل ویژگی‌های سنگ و سیال مخزن را نشان می‌دهند. بنابراین با تحلیل دقیق داده‌ها می‌توان مقادیر لاگ صوتی را تخمین زد. از مهم‌ترین روش‌های تخمین این داده‌ها در گذشته می‌توان به روش رگرسیون چندگانه (Tixier et al., 1975) اشاره کرد. امروزه یکی از بهترین روش‌ها برای پیش‌بینی داده‌های چاه‌های نفت با ضریب اطمینان بالا استفاده از روش‌های یادگیری ماشین<sup>1</sup> (Rajabi et al., 2010; Ramcharitar and Hosein, 2016; Tariq et al., 2017; Zou, 2019; Anemangely et al., 2019; Hosseini et al., 2021) می‌باشد. این روش‌ها با استفاده از ترکیب پارامترهای مختلف دیگر لاگ‌ها، با به حداقل رساندن خطا، پیش‌بینی بهتری نسبت به روش‌های خطی ساده دارد (Akhundi et al., 2014). روش‌های یادگیری ماشین به دو نوع اصلی یادگیری تحت نظارت و بدون نظارت تقسیم می‌شوند. یادگیری نظارت شده در مواردی اعمال می‌شود که در آنها مجموعه‌ای از ورودی‌ها و پاسخ‌های شناخته شده موجود است، درحالی‌که در یادگیری بدون نظارت، پاسخ در دسترس نیست و روش سعی می‌کند الگوهای طبیعی یا خوشه‌ای را در داده‌ها شناسایی کند (McGregor et al., 2004). یادگیری تحت نظارت خود به دو بخش الگوریتم رگرسیون و الگوریتم طبقه‌بندی تقسیم می‌شود. برای تخمین و مدل‌سازی داده‌های پیوسته عددی مانند لاگ‌های پتروفیزیکی از الگوریتم رگرسیون استفاده می‌شود. در این روش، مدل از لاگ‌های مشخص شده بر اساس قرابت با لاگ مورد نظر برای تخمین، به‌عنوان ورودی و خروجی (لاگ انتخاب شده) استفاده می‌کند و لاگ تخمین زده را با لاگ اصلی مقایسه می‌کند. این عمل توسط مدل به‌طور مکرر انجام می‌گیرد تا بهترین لاگ تخمینی با کمترین خطا به دست آید. رگرسیون خطی به‌عنوان یک الگوریتم رایج یادگیری ماشین یک رابطه خطی بین ورودی‌ها و یک پاسخ بر اساس

1. Machine learning

و ۳۰ درصدی برای آموزش مدل یادگیری ماشین و ارزیابی نهایی تقسیم و در ادامه برای یکسان‌سازی تاثیر، نرمال و با استفاده از فرمول‌های ریاضی، داده‌های پرت حذف می‌شوند. در مرحله بعد روش‌های مختلف یادگیری ماشین برای تخمین لاگ صوتی برشی استفاده شده و با استفاده از پارامترهای  $R_2$  و RMSE، بهترین مدل انتخاب و برای کاهش کلی‌گرایی یا جفت‌شدگی زیاد، از هایپریپارامترها استفاده گردید. در ادامه لاگ صوتی برشی تخمین زده شده در چاه B، با داده‌های سرعت برشی حاصل از مغزه در همان چاه مقایسه و مورد ارزیابی قرار می‌گیرد (شکل ۱).

سرعت موج برشی در بخش‌های کنگان و دالان وجود دارد. با این وجود نگار صوتی برشی در این بازه‌های مخزنی اندازه‌گیری نشده است. برای آموزش و یادگیری مدل از اطلاعات چاه شماره A در بخش حاوی نگار سرعت موج برشی استفاده شده است. تست مدل بر اساس نتیجه‌های حاصل از اطلاعات آزمایشگاهی مغزه در چاه شماره B انجام گرفته است.

در ابتدا بر اساس روابط ذاتی بین پارامترهای پتروفیزیکی با سرعت برشی، پارامترهایی که رابطه شناخته شده با لاگ صوتی دارند و همچنین این لاگ‌ها در تمام بخش‌های مورد بررسی وجود دارند انتخاب شدند. داده‌ها به دو بخش ۷۰



شکل ۱. دسته‌بندی مراحل انجام مطالعه

## مدل‌های یادگیری ماشین

در این بخش برای درک بهتر نتیجه‌های حاصله، مدل‌های استفاده شده برای تخمین لاگ صوتی برشی به صورت مختصر توضیح داده شده است.

### رگرسیون خطی<sup>۱</sup>

رگرسیون خطی یک الگوریتم یادگیری ماشین است و ارتباط بین داده‌ها را مشخص می‌کند. دو نوع رگرسیون خطی وجود دارد. نوع اول؛ حالت ابتدایی آن می‌باشد و به آن رگرسیون خطی ساده<sup>۲</sup> (SLR) گفته می‌شود، ارتباط دو

نوع داده مشخص را با هم مقایسه می‌کند. معادله درجه یک حاصل از این روش به ما اجازه می‌دهد در بخش‌هایی که یکی از این داده‌ها وجود نداشته باشد، آن را تخمین زد. به عنوان مثال ارتباط لاگ صوتی برشی و لاگ صوتی فشارشی در چاه‌هایی که هر دو لاگ موجود هستند. نوع دوم، رگرسیون خطی چندگانه<sup>۳</sup> (MLR) است. در این نوع رگرسیون یک

1. near Regression (LR)

2. Simple Linear Regression (SLR)

3. Multiple Linear Regression (MLR)

### رگرسیون بردار پشتیبان<sup>۳</sup>

مدل رگرسیون بردار پشتیبان که به SVR شناخته می‌شود، با استفاده از تابع کرنل<sup>۴</sup> مدل رگرسیونی قوی‌تری را ایجاد می‌کند (Steinwart and Christmann, 2008). تابع کرنل تابعی از فضای ورودی است. در این مقاله از تابع کرنل نوع خطی استفاده شده است. هدف از مدل بردار پشتیبان پیدا کردن خطی می‌باشد که بیشترین فضا (گستره بافر) بین دو مجموعه داده را دارا باشد. داده بعدی که مورد بررسی قرار می‌گیرد در صورت نزدیکی به هر کدام از این دسته‌ها برچسب همان مجموعه را به خود اختصاص خواهد داد. در صورت فاصله‌دار بودن از دو مجموعه داده قبل به عنوان مجموعه سوم شناسایی می‌شود. این مدل برای سه بعدی از حالت خطی خارج می‌شود.

### K-نزدیک‌ترین همسایه<sup>۵</sup>

این الگوریتم با برچسب‌گذاری نقطه‌ها و یافتن نزدیک‌ترین نقطه به آن یک مجموعه با تعداد مشخص (K) را ایجاد می‌کند. به عنوان مثال اگر اندازه K در مدل ۱۰ در نظر گرفته شود، مدل به این تعداد، نزدیک‌ترین نقطه‌های اطراف یک نقطه را در یک مجموعه قرار می‌دهد. در داده‌های عددی، مدل میانگین یا میان اعداد هر مجموعه را به عنوان مقدار کل این مجموعه که خود دیگر یک نقطه جدید است مشخص می‌کند. این مدل علاوه بر پیش‌بینی داده‌ها در خوشه‌بندی داده‌ها نیز بسیار موفق عمل می‌کند.

### شبکه عصبی پرسپترون چند لایه<sup>۶</sup>

این الگوریتم در واقع مجموعه‌ای از شبکه‌های عصبی مصنوعی است و حداقل دارای سه لایه گره شامل: یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است. به جز گره‌های ورودی، هر گره یک نورون است و از یک تابع فعال‌سازی غیرخطی استفاده می‌کند. این رگرسیون با توجه به قدرت تمایز بالای آن برای داده‌هایی بسیار مفید است که به صورت خطی قابل تفکیک نیستند.

رابطه خطی بین چند نوع داده مستقل (به عنوان مثال در اینجا لاگ‌های DTC, NPHI, RHOB & GR) و یک داده وابسته (در اینجا DTS) مشخص می‌شود (معادله ۱) (Kuzmanovski and Aleksovska, 2003).

$$y = a_0 + \sum_{i=1}^n a_i x_i \quad \text{معادله (۱)}$$

x و y به ترتیب داده‌های ورودی (وابسته) و پارامتر خروجی (مستقل)،  $a_1$  ضریب ورودی و  $a_0$  عرض از مبدأ

### رگرسیون جنگل تصادفی<sup>۱</sup>

این روش یادگیری ماشین که توسط بریمان (Breiman, 2001) ارائه شد از تعداد زیادی درخت تصمیم‌گیری استفاده می‌کند تا یک مدل با چندین تصمیم‌گیری ایجاد کند. در مدل درخت تصمیم‌گیری تمامی داده‌ها به صورت یکجا استفاده می‌شوند، به این صورت که داده‌ها به بخش‌های زیادی تقسیم می‌شوند و هر بخش به صورت جداگانه و مستقل با مدل درخت تصمیم‌گیری ارزیابی می‌شوند. در نتیجه مدل، تعداد بسیار زیادی درخت (مدل) آموزش دیده ایجاد می‌شود، به همین علت این روش با عنوان جنگل تصادفی شناخته می‌شود. در نهایت جنگل تصادفی یک مجموعه داده یکسان را در تمامی درخت‌ها اجرا می‌کند. رایج‌ترین پیش‌بینی انجام شده توسط درخت‌ها به عنوان خروجی اصلی انتخاب می‌شود.

### رگرسیون ارتقا یافته<sup>۲</sup>

رگرسیون ارتقا یافته از ترکیب چندین روش برای تصحیح پیش‌بینی خود استفاده می‌کند. به این صورت که در ابتدا با مدل‌های بسیار ساده مانند میانگین داده‌ها، رگرسیون خطی و غیره شروع به پیش‌بینی می‌کند، سپس در هر مرحله با اضافه کردن مدل‌های دیگر، فاصله و خطای داده پیش‌بینی شده را از داده وابسته کاهش می‌دهد. بنابراین مدل نهایی، یک مدل قدرتمند می‌باشد و می‌تواند الگوهایی که بعضی از مدل‌ها نمی‌توانند تشخیص دهند، بررسی و پیدا کند. در نتیجه در مطالعاتی که تعداد داده‌ها بسیار زیاد هستند و روابط غیرخطی دارند کارایی بالایی دارد.

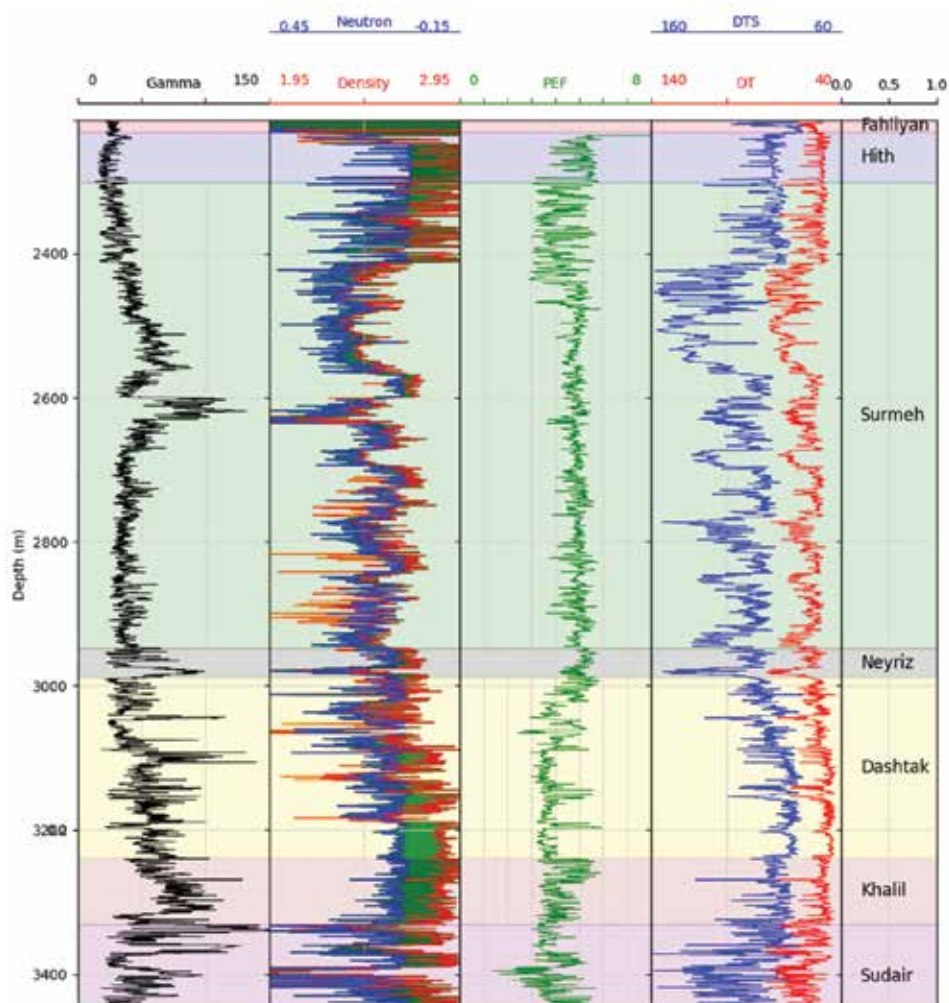
1. Random forest  
2. Gradient boosting regressor  
3. Support vector regressor  
4. Kernel  
5. K neighbors regressor  
6. MLP Regressor

## بحث

### آماده‌سازی داده‌ها

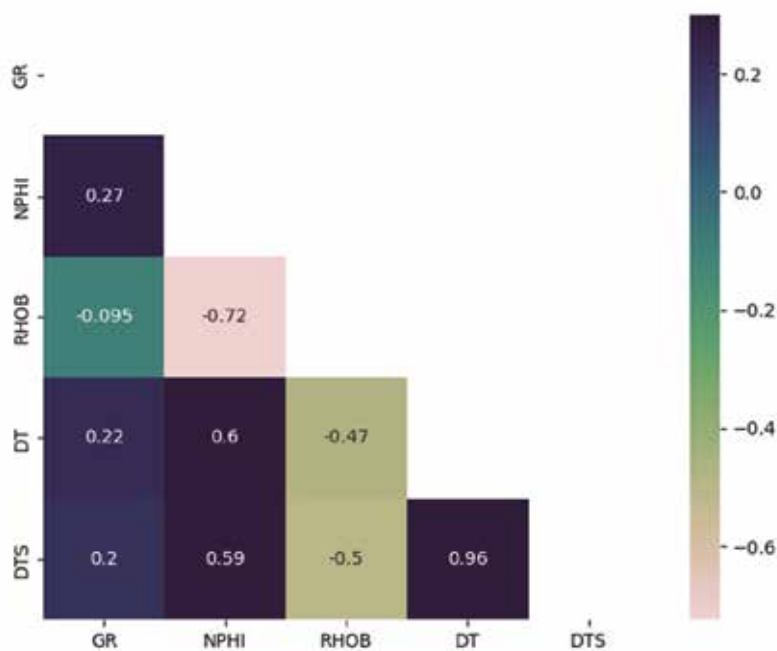
(لاگ صوتی برشی) قرابت داشته باشند. برای این هدف از ضریب همبستگی (CC) استفاده شده و همیشه بین منفی یک و مثبت یک می‌باشد به صورتی که هر مقدار به منفی یک نزدیک‌تر باشند، نشان‌دهنده رابطه معکوس قوی‌تر و هر مقدار به مثبت یک نزدیک باشند نشان‌دهنده رابطه مستقیم قوی‌تری می‌باشند. ضریب همبستگی صفر نیز نداشتن ارتباط بین داده‌ها را نشان می‌دهد (شکل ۳). در نهایت، لاگ‌های انتخاب شده برای استفاده در مدل‌های یادگیری ماشین، شامل لاگ‌های گاما، نوترون، دانسیته، صوتی فشارشی به‌عنوان ورودی و صوتی برشی به‌عنوان خروجی هدف می‌باشند.

برای به دست آوردن مدل پیش‌بینی قابل اعتماد، پیش‌پردازش داده‌ها برای شناسایی بازه‌ها و لاگ‌های مورد استفاده دارای اهمیت می‌باشد. در این میدان تنها یک چاه دارای لاگ صوتی برشی (۱۴۰۰ متر) می‌باشد (بخش بالایی چاه شماره A). داده‌ها در قالب فایل LAS مورد استفاده قرار گرفتند. برای انتخاب لاگ، باید لاگ‌هایی انتخاب شوند که بیشترین تاثیر ذاتی را در مقادیر لاگ صوتی برشی داشته باشند و همچنین این لاگ‌ها در تمامی چاه‌های آموزشی و هدف وجود داشته باشند. بالا و پایین لاگ‌های مختلف مشخص شد و بازه‌های بدون داده حذف شد (شکل ۲). همچنین باید این لاگ‌ها از منظر ریاضیاتی با لاگ هدف



شکل ۲. لاگ‌های انتخاب شده برای آموزش مدل‌های یادگیری ماشین

تخمین لاگ صوتی برشی با استفاده از روش‌های یادگیری ماشین، و مقایسه با داده‌های حاصل از مغزه ...



شکل ۳. ضریب همبستگی بین لاگ‌های موجود

کمتر باشد، روش استفاده شده می‌تواند منجر به نتیجه‌های تخمین لاگ بهتری شود. با توجه به تعداد داده‌های حذف شده که در جدول ۱ آورده شده است، روش‌های انحراف معیار و حداقل کوواریانس دارای کمترین داده حذف شده هستند. در ادامه، نمودارهای باکس پلات رسم شد (شکل ۴). با توجه به تعداد داده‌های حذف شده، نمودار باکس پلات رسم شده و همچنین تخمین اولیه لاگ صوتی، در این مقاله برای نرمال‌سازی داده‌ها از روش انحراف معیار استفاده شده است.

برای حذف تاثیر اندازه یا واحد داده‌ها و یکسان‌سازی آنها، روش‌های متفاوتی برای نرمال‌سازی مورد استفاده قرار گرفت. هدف از بررسی این روش‌ها انتخاب بهترین روش ریاضیاتی برای نرمال کردن داده‌ها می‌باشد. روش‌های مورد استفاده شامل انحراف معیار، جنگل انزوا، حداقل کوواریانس و فاکتور پرتی می‌باشند. برای انتخاب روش نرمال‌سازی دو ویژگی مورد بررسی قرار گرفت. هر یک از این روش‌ها برای نرمال‌سازی داده‌ها نیاز به حذف داده‌های پرت دارند. بنابراین هر مقدار که تعداد این داده‌های حذف شده

جدول ۱. تعداد داده‌های باقیمانده از لاگ‌های انتخابی بعد از اعمال روش‌های نرم‌سازی مختلف

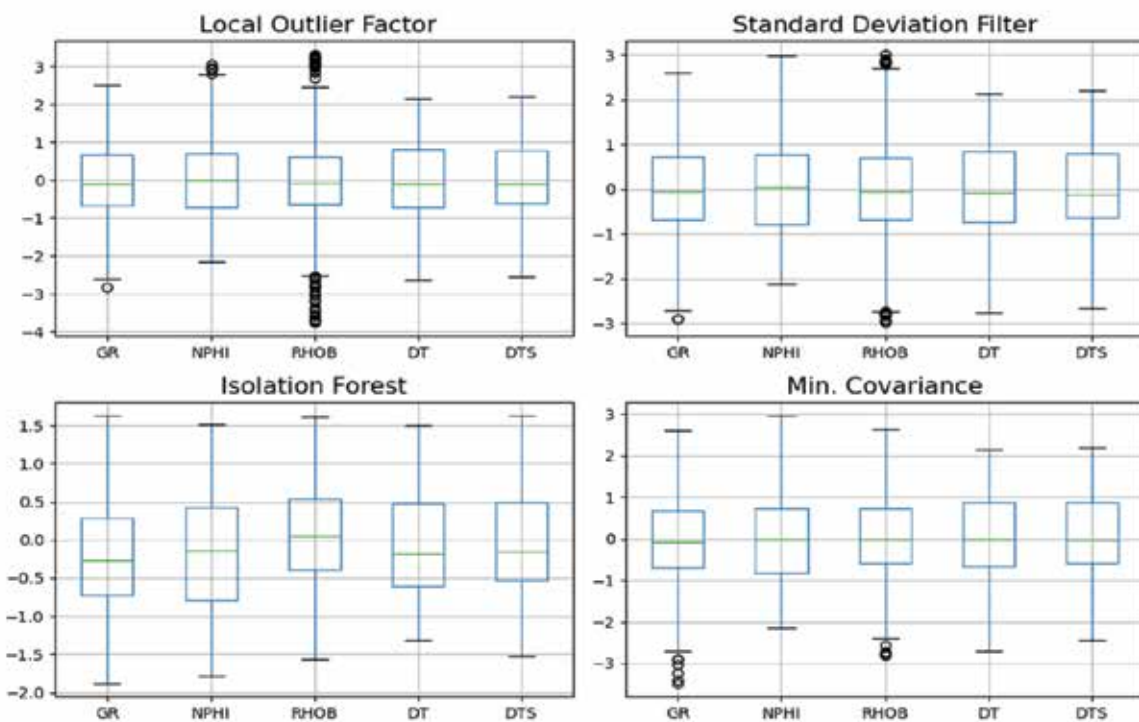
لاگ‌های اولیه	Standard Deviation	Min. Covariance	Isolation Forest	Outlier Factor
4003	3984	3602	2002	2802

## توسعه مدل

خطا<sup>۳</sup> (معادله ۴) نیز برای تمامی مدل‌ها محاسبه و بهترین مدل برای ادامه کار انتخاب شد (جدول ۲). مقدار اندیس ارزیابی بین صفر و یک می‌باشد و هر مقدار که عدد حاصله

برای تعیین درستی عملکرد مدل‌های معرفی شده، داده‌های حاوی لاگ صوتی برشی به دو بخش تقسیم شدند. بخش اول شامل ۷۰ درصد داده‌ها با هدف آموزش مدل و بخش ۳۰ درصدی با هدف تست مدل آموزش دیده تقسیم شد. همچنین مقادیر اندیس ارزیابی<sup>۱</sup> ( $R^2$ ) (معادله ۲)، میانگین مربعات خطا<sup>۳</sup> (معادله ۳) و ریشه میانگین مربع‌های

1. Coefficient of determination  
2. Mean squared error (MSE)  
3. RMSE



شکل ۴. نمودار باکس پلات داده‌ها بعد از نرمال‌سازی

موجود و لاگ صوتی برشی تخمین زده شده را برای تمامی روش‌ها نشان داده است. همچنین مقادیر اندیس ارزیابی و ریشه میانگین مربعات خطا برای تمامی روش‌ها محاسبه شد (جدول ۲).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n y_i^2 - \frac{\sum_{i=1}^n y_i'^2}{n}} \quad \text{معادله (۲)}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad \text{معادله (۳)}$$

$$RMSE = \sqrt{MSE} \quad \text{معادله (۴)}$$

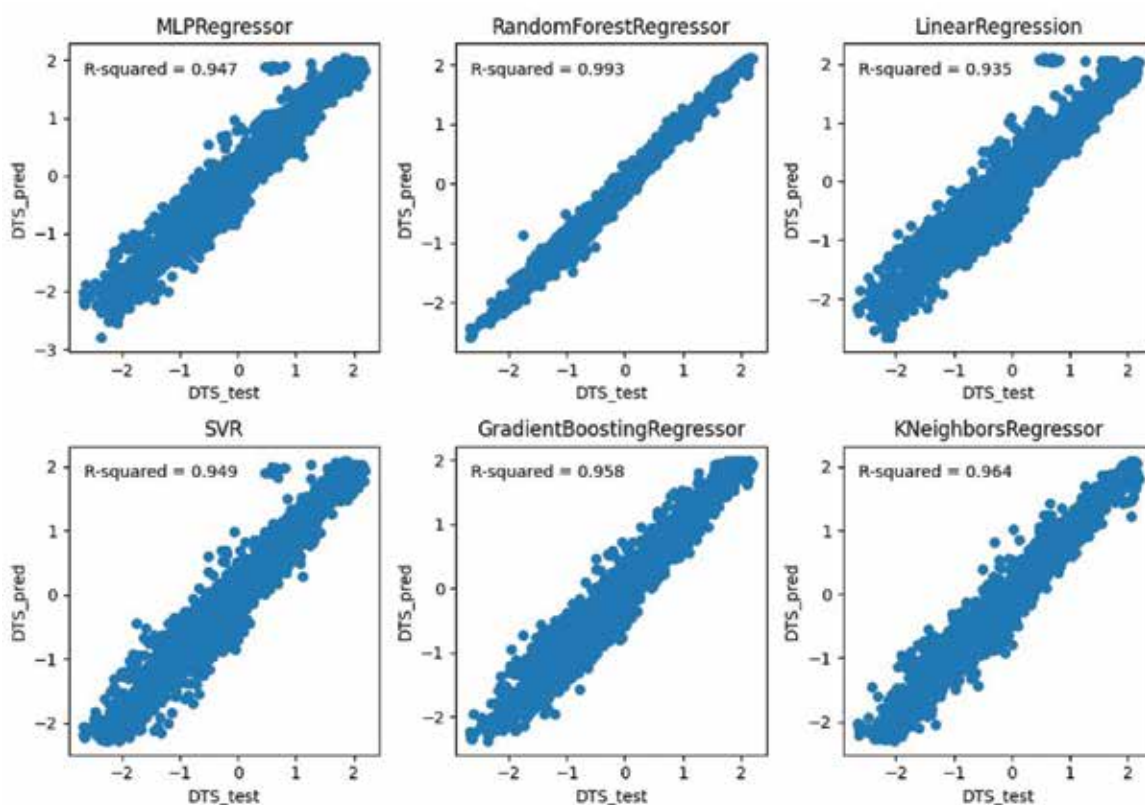
$y_i$  مقدار اندازه‌گیری شده،  $y'_i$  مقدار پیش‌بینی شده

به یک نزدیک‌تر باشد نشان‌دهنده همبستگی بالاتر داده‌ها می‌باشد. میانگین مربع‌های خطا تفاوت بین مقدار اصلی و مقدار تخمین زده شده را برآورد می‌کند و هرچه این مقدار به صفر نزدیک‌تر باشد، نشان‌دهنده میزان خطای محاسباتی کمتر می‌باشد. ولی ریشه میانگین مربعات خطا رابطه معکوسی با همبستگی دارد، به این صورت که مقادیر پایین نشان‌دهنده همبستگی بالاتر هستند. بر این اساس روش‌های مختلف یادگیری ماشین معرفی شده در بخش قبلی مورد استفاده قرار گرفت و لاگ صوتی برشی تخمین زده شده است. شکل ۵ نمودار پلات لاگ صوتی برشی

جدول ۲. مقادیر اندیس ارزیابی و ریشه میانگین مربعات خطا برای مدل‌های مختلف

Index	MLP Regressor	Random Forest Regressor	Linear Regression	SVR	Gradient Boosting Regressor	K Neighbors Regressor
R-squared	۰٫۹۵	۰٫۹۹	۰٫۹۴	۰٫۹۵	۰٫۹۶	۰٫۹۶
RMSE	۰٫۲۲۸	۰٫۰۸۷	۰٫۲۵۶	۰٫۲۲۶	۰٫۲۰۵	۰٫۱۸۸
MSE	۰٫۰۵	۰٫۰۱	۰٫۰۷	۰٫۰۵	۰٫۰۴	۰٫۰۴

تخمین لاگ صوتی برشی با استفاده از روش‌های یادگیری ماشین، و مقایسه با داده‌های حاصل از مغزه ...



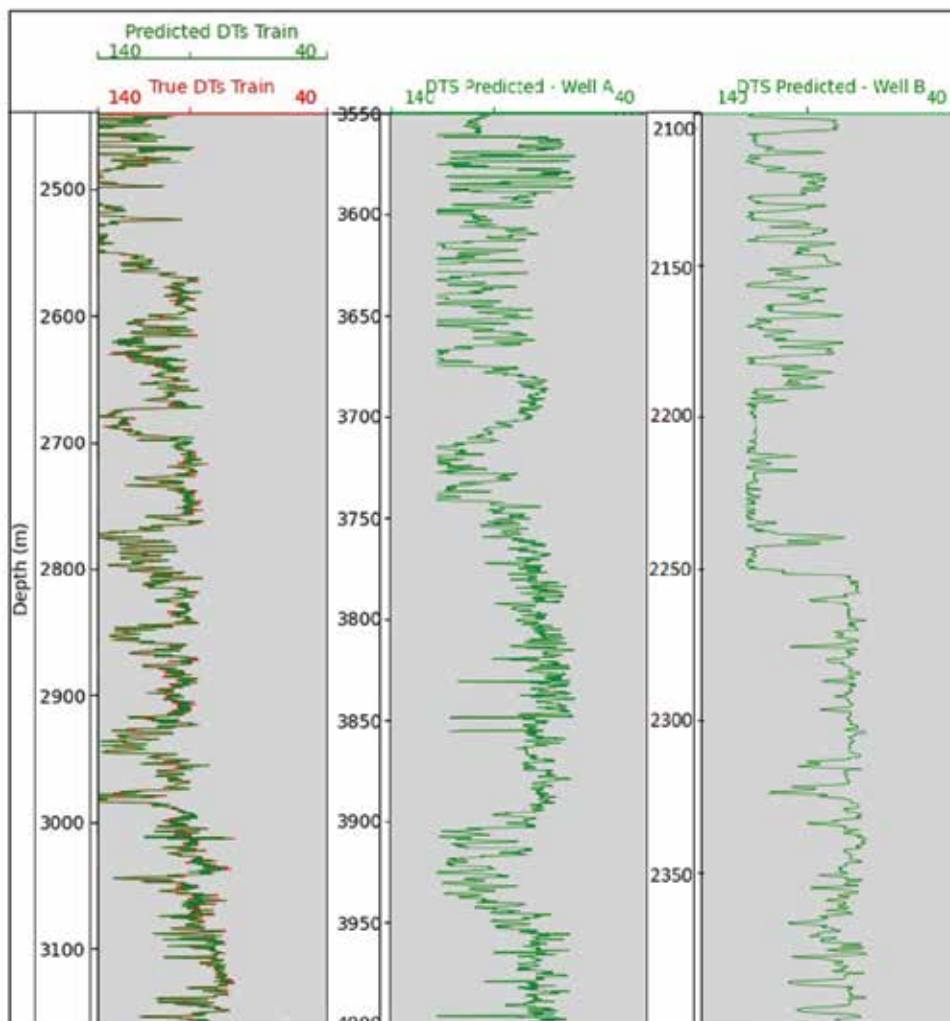
شکل ۵. نمودار کراس پلات لاگ‌های صوتی برشی تخمین زده شده از روش‌های یادگیری ماشین در مقابل داده‌های لاگ صوتی برشی اصلی (ثبت شده در چاه)

هایپیرامترها از تابع GridSearchCV استفاده شد. این تابع با جستجوی گسترده در شبکه داده‌ها، بهترین پارامترها را برای مدل انتخاب می‌کند. هایپیرامترهای به دست آمده از تابع GridSearchCV شامل 'max\_depth': 40, 'min\_samples\_split': 5 و 'min\_samples\_leaf': 5 می‌باشند. اندیس ارزیابی در این حالت برابر ۰/۹۸ RMSE و برابر ۰/۱۸ می‌باشد. بنابراین از این مدل نهایی برای تخمین لاگ‌های صوتی برشی استفاده شد. شکل ۵ سمت چپ لاگ صوتی برشی تخمین زده شده در بخش داده‌های تست در کنار لاگ صوتی برشی اولیه همان بازه را نشان می‌دهد. همچنین در این شکل (۶) لاگ‌های صوتی برشی تخمین زده شده سازندهای کنگان-دالان در چاه A و B نشان داده شده است.

بر اساس نتیجه‌های حاصل در جدول شماره ۲ مدل رگرسیون جنگل تصادفی دارای بالاترین مقدار اندیس ارزیابی (۰/۹۹) و کمترین مقدار ریشه میانگین مربعات خطا (۰/۰۹)، و رگرسیون خطی چندگانه دارای کمترین مقدار اندیس ارزیابی (۰/۹۴) با بیشترین مقدار برای ریشه میانگین مربعات خطا (۰/۲۶) بود. بنابراین مدل جنگل تصادفی برای ادامه محاسبات انتخاب شد. مدل انتخاب شده در بخش داده‌های تست نیز با دقت ۹۸ درصد لاگ صوتی برشی را تخمین زد و نشان دهنده دقت بالای مدل می‌باشد. در ادامه برای کاهش حداقل خطا و دوری از خطای کلیت بخشی<sup>۱</sup> یا بیش برآزش<sup>۲</sup> داده‌ها که در زمان خیلی پیچیده بودن یا ساده بودن مدل به وجود می‌آید، از هایپیرامترها برای بهینه کردن مدل استفاده شد. برای تعیین مقادیر بهینه

1. Generalization  
2. Overfitting





شکل ۶. لاگ‌های صوتی برشی تخمین زده شده در بخش آموزش چاه A، مقایسه لاگ صوتی برشی در داده‌های تست با لاگ صوتی برشی تخمین زده شده در همان بازه و لاگ‌های تخمین زده شده سازندهای دالان و کنگان در چاه‌های A و B

### مقایسه لاگ تخمین زده شده با داده‌های مغزه

برشی نیاز است، این داده‌ها به واحد میکروثانیه بر فوت برگردانده شوند. بعد از انجام این تغییرات داده‌های حاصل از مغزه بر روی لاگ صوتی برشی تخمین زده شده پلات شد (شکل ۷). شکل ۷ نشان می‌دهد روش مورد استفاده (تخمین لاگ صوتی برشی بر اساس مدل جنگل تصادفی)، کارایی بسیار خوبی برای تخمین لاگ صوتی برشی از خود نشان می‌دهد. توزیع دوبعدی داده‌های حاصل از مغزه و داده‌های تخمین زده شده مقدار اندیس ارزیابی ۰/۹۶ را نشان می‌دهد، با توجه به تعداد کم داده‌های مغزه، نتیجه

برای صحت‌سنجی داده‌های لاگ‌های صوتی برشی به‌دست آمده در چاه‌های جدید، نیاز به مقایسه با داده‌های صوتی حاصل از مغزه در همان بازه می‌باشد. برای این منظور پارامترهای سرعت مغزه‌ها با استفاده از روش تست التراسونیک<sup>۱</sup> محاسبه شده‌اند (جدول ۳). در این مطالعه از داده‌های ۱۰ پلاگ برای اندازه‌گیری سرعت برشی و فشارشی با ضخامت ۲۲ میلی‌متر با استاندارد ASTM-D2845-83 (1985) برای ارزیابی سرعت موج‌های الاستیک مورد استفاده قرار گرفته است. واحد داده‌های سرعت موج برشی حاصله، متر بر ثانیه می‌باشد. بنابراین برای مقایسه با داده‌های صوتی

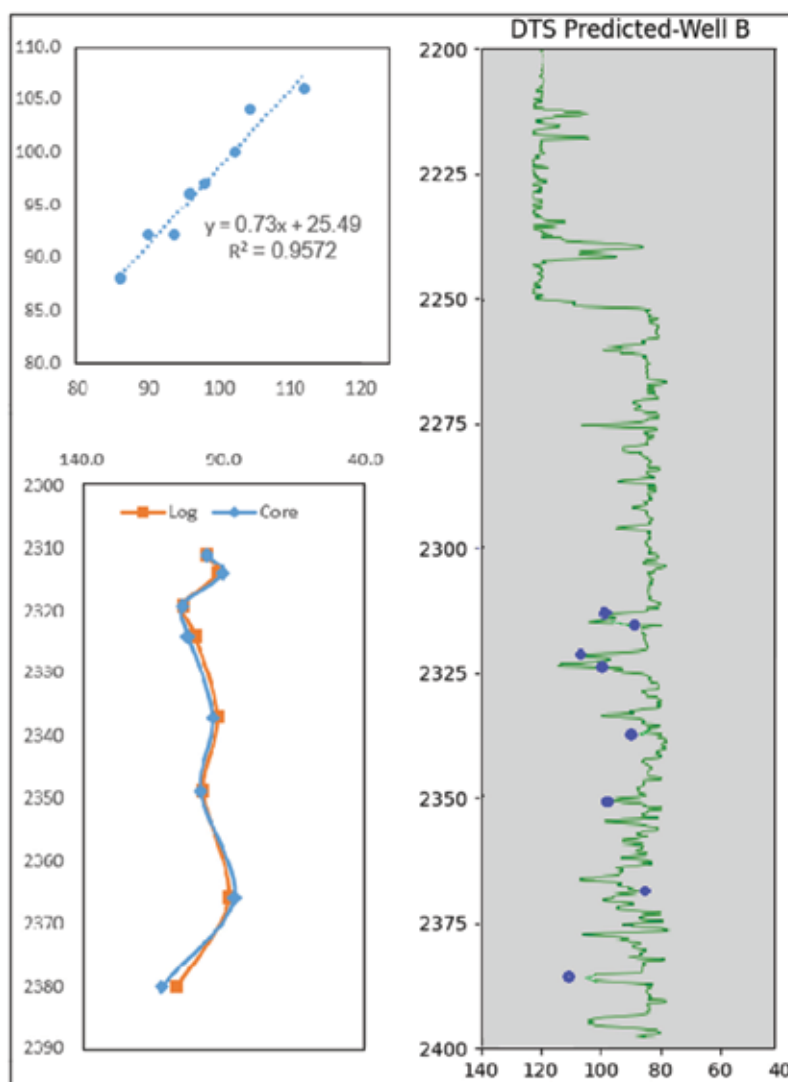
1. Ultrasonic

تخمین لاگ صوتی برشی با استفاده از روش‌های یادگیری ماشین، و مقایسه با داده‌های حاصل از مغزه ...

قابل قبولی می‌باشد. بنابراین می‌توان از این روش در تمامی استفاده کرد و لاگ صوتی برشی را تخمین زد. چاه‌های این میدان که داده لاگ صوتی برشی وجود ندارد،

جدول ۳. داده‌های سرعت موج برشی و فشارشی حاصله از مغزه در سازندهای دالان و کنگان

Sample No.	Well	Depth	Length (mm)	Density (g/cc)	Time ( $\mu$ s)	Vp (m/s)	Vs (m/s)	DTs ( $\mu$ s / f)	DTs-esti ( $\mu$ s / f)
۱	B	۲۳۱۱	۸۶/۵	۲/۹۷	۱۴/۵	۵۹۶۶	۳۱۷۸	۹۶	۹۶/۰
۲	B	۲۳۱۴	۸۸	۲/۹۸	۱۴/۷	۵۹۸۶	۳۳۸۹	۹۰	۹۲/۰
۳	B	۲۳۱۹	۹۱	۲/۹۵	۱۷/۸	۵۱۱۲	۲۹۱۹	۱۰۴	۱۰۴/۰
۴	B	۲۳۲۴	۸۱/۵	۲/۹۲	۱۶	۵۰۹۴	۲۹۸۰	۱۰۲	۱۰۰/۰
۵	B	۲۳۳۷	۸۶	۲/۹۴	۱۴/۱	۶۰۹۹	۳۲۵۱	۹۴	۹۲/۰
۶	B	۲۳۴۹	۹۲	۲/۷۲	۲۰/۹	۴۴۰۲	۳۱۰۲	۹۸	۹۷/۰
۷	B	۲۳۶۶	۸۶	۲/۸۹	۱۳/۴	۶۴۱۸	۳۵۴۰	۸۶	۸۸/۰
۸	B	۲۳۸۰	۹۰	۲/۶۹	۲۰/۳	۴۴۳۳	۲۷۱۱	۱۱۲	۱۰۶/۰



شکل ۷. A) توزیع دوبعدی زمان عبور موج برشی تخمین زده شده با نتیجه‌های اندازه‌گیری شده آزمایشگاهی، B و C) پلات داده‌های صوتی برشی حاصل از مغزه بر روی لاگ صوتی برشی تخمین زده شده در چاه B

Science and Engineering, 175, 407-429. Doi: <https://doi.org/10.1016/j.petrol.2018.12.054>

- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5-32. Doi: <https://doi.org/10.1023/A:1010933404324>

- Eskandari, H., Rezaee, M.R. and Mohammadnia, M., 2004. Application of multiple regression and artificial neural network techniques to predict shear wave velocity from wireline log data for a carbonate reservoir South-West Iran. *CSEG recorder*, 42, 40-48.

- Fjaer, E., Holt, R.M., Horsrud, P. and Raaen, A.M., 2008. *Petroleum Related Rock Mechanics*. Elsevier Science Publisher: Amsterdam, The Netherlands.

- Forkuor, G., Hounkpatin, O.K., Welp, G. and Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1), p.e0170478. Doi: <https://doi.org/10.1371/journal.pone.0170478>

- Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298), 369-375. Doi: <https://doi.org/10.1080/01621459.1962.10480665>

- Heiat, A., 2002. Comparison of artificial neural network and regression models for estimating software development effort. *Information and software Technology*, 44(15), 911-922. Doi: [https://doi.org/10.1016/S0950-5849\(02\)00128-3](https://doi.org/10.1016/S0950-5849(02)00128-3)

- McGregor, A., Hall, M., Lorier, P. and Brunskill, J., 2004. Flow clustering using machine learning techniques. In *Passive and Active Network Measurement: 5th International Workshop, PAM 2004, Antibes Juan-les-Pins, France, April 19-20, 2004*. Proceedings 5, 205-214. Springer Berlin Heidelberg.

- Rajabi, M., Bohlooli, B. and Ahangar, E.G., 2010. Intelligent approaches for prediction

## نتیجه‌گیری

لاگ صوتی برشی به دلیل هزینه بالا و زمان نمودارگیری در چاه‌های محدودی گرفته می‌شود. این لاگ در تفاسیر پتروفیزیکی، ساخت مدل‌های پایداری دیواره چاه و مدل‌سازی‌های ژئومکانیکی از اهمیت بالایی برخوردار است. مطالعه حاضر به تخمین لاگ صوتی برشی از لاگ‌های پتروفیزیکی معمول با استفاده از روش‌های یادگیری ماشین پرداخته است. آماده‌سازی و نرمال‌سازی داده‌ها در انجام روش‌های یادگیری ماشین اهمیت بالایی دارد. در این مطالعه برای این منظور از روش‌های مختلفی استفاده شد. نرمال‌سازی با روش انحراف معیار بهترین نتیجه‌ها با کمترین داده حذف شده را نشان داد. این مطالعه نشان داد برای تخمین لاگ صوتی برشی در چاه‌های نفت، استفاده از مدل‌های یادگیری ماشین بر پایه تصمیم‌گیری مانند جنگل تصادفی بهترین نتیجه‌ها را خواهد داد. مقایسه نتیجه‌ها در این مطالعه نشان داد روش رگرسیون جنگل تصادفی کارایی بهتری نسبت به دیگر روش‌های مورد استفاده دارد (با قدرت تخمین بالای ۹۸ درصد). بنابراین از این روش برای محاسبات استفاده شد. در این روش اندیس ارزیابی برای داده‌های آموزش و همچنین داده‌های تست ۹۸ درصد بود. همچنین برای صحت‌سنجی، لاگ‌های تخمین زده شده با داده‌های سرعت حاصل از مغزه مقایسه شد و نتایج نشان‌دهنده قرابت بالای این داده‌ها ( $R^2=0.96$ ) می‌باشد.

## منابع

- Akhundi, H., Ghafouri, M. and Lashkaripour, G.R., 2014. Prediction of shear wave velocity using artificial neural network technique, multiple regression and petrophysical data: A case study in Asmari reservoir (SW Iran). *Open Journal of Geology*, 4, 303-313. Doi: <https://doi.org/10.4236/ojg.2014.47023>

- Anemangely, M., Ramezanzadeh, A. and Behboud, M.M., 2019. Geomechanical parameter estimation from mechanical specific energy using artificial intelligence. *Journal of Petroleum*

- of compressional, shear and Stoneley wave velocities from conventional well log data: A case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran). *Computers & Geosciences*, 36(5), 647-664. Doi: <https://doi.org/10.1016/j.cageo.2009.09.008>
- Ramcharitar, K. and Hosein, R., 2016, June. Rock Mechanical Properties of Shallow Unconsolidated Sandstone Formations. Paper presented at the SPE Trinidad and Tobago Section Energy Resources Conference, Port of Spain, Trinidad and Tobago. Doi: <https://doi.org/10.2118/180803-MS>
- Hosseini, Z., Gharechelou, S., Mahboubi, A., Moussavi-Harami, R., Kadkhodaie-Ilkhchi, A., and Zeinali, M., 2021, Shear wave velocity estimation utilizing statistical and multi-intelligent models from petrophysical data in a mixed carbonate-siliciclastic reservoir, SW Iran. *Iranian Journal of Oil and Gas Science and Technology*, 10(1), 15-39. doi: <https://10.22050/ijogst.2020.241095.1556>
- Steinwart, I. and Christmann, A., 2008. *Support vector machines*. Springer Science and Business Media.
- Tariq, Z., Elkatatny, S.M., Mahmoud, M.A., Abdulraheem, A., Abdelwahab, A.Z. and Woldeamanuel, M., 2017, June. Estimation of Rock Mechanical Parameters Using Artificial Intelligence Tools. In *ARMA US Rock Mechanics/Geomechanics Symposium (ARMA-2017)*. ARMA.
- Tixier, M.P., Loveless, G.W. and Anderson, R.A., 1975. Estimation of formation strength from the mechanical-properties log (includes associated paper 6400). *Journal of Petroleum Technology*, 27(03), 283-293. Doi: <https://doi.org/10.2118/4532-PA>
- Xu, Y., Zhang, H. and Guan, Z., 2021. Dynamic characteristics of downhole bit load and analysis of conversion efficiency of drill string vibration energy. *Energies*, 14(1), 229. Doi: <https://doi.org/10.3390/en14010229>
- Zou, X., 2019. Application of machine learning in shear wave prediction of jiaoshiba shale gas horizontal well. *Jiangnan Petroleum Science and Technology*, 29(4), 16-22.
- Kuzmanovski, I. and Aleksovska, S., 2003. Optimization of artificial neural networks for prediction of the unit cell parameters in orthorhombic perovskites. Comparison with multiple linear regression. *Chemometrics and Intelligent Laboratory Systems*, 67(2), pp.167-174.