

## Comparison of artificial intelligence algorithms to predict pore pressure using petrophysical log data

Meysam Rajabi<sup>\*1</sup>, Hamzeh Ghorbani<sup>2</sup>, Sahar Lajmorak<sup>3</sup>

1- Assistant Professor, Department of Mining Engineering, Birjand University of Technology, Birjand, Iran

2- Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

3- Master of Geology, Faculty of Earth Sciences, Zanjan University of Basic Sciences, Iran

### ABSTRACT

Pore pressure is one of the most important reservoir-drilling parameters and knowledge of this pressure is essential for drilling costs, well safety and prevention of potential hazards. Research has shown that experimental equations have good performance accuracy only for certain regions. Most of these experimental equations have been compiled and developed based on a limited data set. Therefore, these correlations are valid in the range of changes in the parameters of those fields and are not valid for other areas. Therefore, artificial intelligent methods have given way to empirical equations. In this study, 2827 data related to three wells from one of the oil fields located in the southwest of Iran have been used. The input variables used in this paper to predict the pore pressure include 9 variables that have been selected using the feature selection method. In this study, 4 artificial intelligence algorithms include; random forest algorithm, support vector regression algorithm, artificial neural network algorithm and decision tree algorithm have been used to predict the pore pressure. After reviewing the results, it was found that the performance accuracy of the decision tree algorithm is higher than the other three algorithms (performance accuracy for the entire data set including  $R^2 = 0.9985$  and  $RMSE = 14.460$  psi). Among the advantages of this algorithm compared to other algorithms are the best results without the need for statistical knowledge, separation of unnecessary data, short time to prepare data and reduction of relative error by finding the main node of the decision maker and analyzing it. Therefore, it can be concluded that with the development of this technique, it is possible to have high performance accuracy for a small amount of data from each field.

### ARTICLE INFO

**Receive Date:** 20 October 2021

**Revise Date:** 06 March 2022

**Accept Date:** 05 May 2022

### Keywords:

Pore pressure  
Artificial intelligence  
algorithms  
Petrophysical data  
Decision tree algorithm  
Feature selection

All rights reserved to Iranian Society of Structural Engineering.

doi: <https://doi.org/10.22065/jsce.2022.309523.2600>

\*Corresponding author: Meysam Rajabi.  
Email address: m.rajabi@birjandut.ac.ir

## پیش بینی فشار منفذی داده‌های گمانه‌های پتروفیزیکی با بهره‌گیری از روش‌های مبتنی بر هوش مصنوعی

میثم رجبی<sup>۱\*</sup>، حمزه قربانی<sup>۲</sup>، سحر لجم اورک<sup>۳</sup>

۱- استادیار گروه مهندسی معدن، دانشگاه صنعتی بیرجند، بیرجند، ایران

۲- باشگاه پژوهشگران جوان و نخبگان، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران

۳- کارشناسی ارشد زمین‌شناسی، دانشکده علوم زمین، دانشگاه تحصیلات تکمیلی علوم پایه زنجان، ایران

### چکیده

فشار منفذی یکی از مهم‌ترین پارامترهای مخزنی-حفاری به شمار می‌رود و آگاهی از این فشار، برای کاهش هزینه‌های حفاری، افزایش ایمنی چاه و پیشگیری از خطرات احتمالی ضروری است. امروزه روش‌های هوشمند جایگزین خوبی برای معادلات تجربی هستند که تنها بر اساس مجموعه داده‌های محدود، جمع‌آوری و توسعه یافته بودند. مطالعه پیش رو، از ۲۸۲۷ مجموعه داده‌ای مربوط به سه چاه از میدان‌ها نفتی واقع در جنوب غرب ایران بهره برده است. متغیرهای ورودی مورد استفاده به منظور پیش‌بینی فشار منفذی شامل ۹ متغیر بوده که با استفاده از روش انتخاب ویژگی، برگزیده شده‌اند. در این مطالعه از چهار الگوریتم هوش مصنوعی شامل الگوریتم جنگل تصادفی<sup>۱</sup>، الگوریتم رگرسیون بردار پشتیبان<sup>۲</sup>، الگوریتم شبکه عصبی مصنوعی<sup>۳</sup> و الگوریتم درخت تصمیم<sup>۴</sup> به منظور پیش‌بینی فشار منفذی استفاده شده است. نتایج نشان می‌دهد که دقت عملکرد الگوریتم درخت تصمیم، بیشتر از سه الگوریتم دیگر است. مقدار مربع ضریب همبستگی<sup>۵</sup> ( $R^2$ ) و مجذور میانگین مربعات خطا<sup>۶</sup> ( $RMSE$ ) برای تخمین فشار منفذی توسط الگوریتم درخت تصمیم به ترتیب ۰/۹۹۸۵ و ۱۴/۴۶۰ psi برآورد شد. از جمله مزیت‌های این الگوریتم، ارائه بهترین نتیجه بدون نیاز به دانش آماری، جدا کردن داده‌های غیرضروری، آماده‌سازی داده‌ها در زمان کوتاه و کاهش خطای نسبی با یافتن گره اصلی تصمیم گیر است. بنابراین می‌توان چنین تحلیل نمود که با توسعه این روش هوشمند، برای تعداد داده‌های کم از هر میدان، دقت عملکرد بالایی نتیجه می‌شود.

کلمات کلیدی: فشار منفذی، الگوریتم‌های هوش مصنوعی، داده‌های پتروفیزیکی، الگوریتم درخت تصمیم، انتخاب ویژگی.

شناسه دیجیتال:		سابقه مقاله:				
doi:	https://doi.org/10.22065/jsce.2022.309523.2600	چاپ	انتشار آنلاین	پذیرش	بازنگری	دریافت
	10.22065/jsce.2022.309523.2600	۱۴۰۱/۱۱/۳۰	۱۴۰۱/۰۲/۱۵	۱۴۰۱/۰۲/۱۵	۱۴۰۰/۱۲/۱۵	۱۴۰۰/۰۷/۲۸
				*نویسنده مسئول:		
				پست الکترونیکی:		
				میثم رجبی		
				m.rajabi@birjandut.ac.ir		

<sup>1</sup> Random forest algorithm

<sup>2</sup> Support vector regression algorithm

<sup>3</sup> Artificial neural network algorithm

<sup>4</sup> Decision tree

<sup>5</sup> Correlation coefficient

<sup>6</sup> Root mean square error

## ۱- مقدمه

فشار منفذی یکی از مهم‌ترین پارامترهای تأثیرگذار در فرایند حفاری و اکتشاف از جمله بازده حفاری و هزینه عملیات چاه مانند طراحی چاه، تجزیه و تحلیل پایداری چاه، طراحی پوشش، طراحی برنامه گل، عملیات حفاری و بهینه‌سازی ساختار است [۱-۴]. تعیین دقیق فشار منفذی کمک شایانی به تولید و تزریق انتخابی و ردیابی مسیرهای نشت هیدروکربن‌ها کرده و از هدر رفتن گل حفاری جلوگیری می‌کند [۵، ۶]. تخمین اشتباه این پارامتر مهم سبب ایجاد مشکلات عدیده‌ای در صنعت نفت و گاز همچون از دست دادن گل حفاری که نتیجه آن فوران چاه و آسیب جبران‌ناپذیر مالی (برای صنعت نفت) و جانی (برای کارکنان حفاری) می‌شود [۷].

فشار منفذی یا فشار سازندی، فشار سیالات درون منفذ سازند بوده که ناشی از پتانسیل هیدرولیکی است [۸]. این فشار زمانی به‌عنوان یک فشار ایمن شناخته می‌شود که فشار سیال حفاری بیشتر از این فشار و کمتر از فشار شکست سازند باشد [۹] در غیر اینصورت مشکلات و معضلات عدیده‌ای برای عملیات چاه به وجود می‌آید [۱۰].

در طول سالیان اخیر محققان بسیاری از روش‌های هوش مصنوعی به منظور پیش‌بینی پارامترهای مختلف استفاده کردند [۱۱]. یکی از مهم‌ترین پارامترهای مهم مخزنی که با استفاده از هوش مصنوعی تخمین زده شد، تعیین فشار منفذی می‌باشد. اولین مطالعاتی که به‌منظور پیش‌بینی فشار منفذی صورت گرفت و منجر به ارائه یک معادله‌ی تجربی گردید توسط ترزاقی<sup>۷</sup> در سال ۱۹۴۳ ارائه شد. بر اساس این تئوری فشار سرپاره به وسیله فشار منفذی، خنثی‌شده و تمامی اثرات تنش از جمله تغییرات سرعت موج الاستیک را به‌عنوان تنش مؤثر در نظر گرفت. تحقیقات هاتمن و جانسون<sup>۸</sup> در سال ۱۹۶۵، با توجه به خواص شیل‌ها و بر اساس انحراف خط نرمال در اندازه‌گیری سرعت صوت در نگاره انجام شد [۱۳]. در همان سال بینگهام نیز معادله‌ی (۱) را به‌منظور پیش‌بینی فشار غیرطبیعی سازند بر اساس پارامترهای حفاری ارائه داد [۱۴]:

$$ROP = d * RPM^E * \left(\frac{WOB}{D}\right)^{D_e} \quad (1)$$

که در آن؛  $ROP$  نرخ نفوذ؛  $d$  ثابت حفاری؛  $RPM$  سرعت چرخش حفاری؛  $E$  ثابت سرعت چرخش؛  $D$  قطر مته؛  $WOB$  وزن روی مته و  $D_e$  وزن روی مته توانی است.

در سال ۱۹۶۶ جردن و شرلی<sup>۹</sup> با تعریف ترم جدیدی به نام  $De$  معادله‌ی بینگهام را تغییر داده (معادله ۲) و از آن به‌منظور تغییر فشار معمولی به غیرطبیعی استفاده کردند (در صورتی که دانسیته گل ثابت نگه داشته شود) [۱۵].

$$D_e = \left( \frac{\log\left(\frac{ROP}{RPM}\right)}{\log\left(\frac{12WOB}{10^6 D}\right)} \right) \quad (2)$$

ترزاقی و همکاران در سال ۱۹۹۶، فشار اضافی در یک نقطه را که توسط ماتریس سنگ و سیال موجود در ماتریس سنگ ایجاد می‌شود با معادله (۳) گزارش کرده‌اند [۱۶]. میزان  $S_{over}$  تابعی از چگالی سنگ است و از طریق معادله‌ی (۴) محاسبه می‌شود.

$$S_{over} = S_{eff} + PP \quad (3)$$

$$S_{over} = 0.433 \int_0^h \rho gh dh \quad (4)$$

در معادلات بالا؛  $S_{over}$  فشار سرپاره؛  $S_{eff}$  فشار مؤثر؛  $PP$  فشار منفذی؛  $\rho$  دانسیته سنگ؛  $g$  شتاب گرانشی؛  $h$  عمق است.

<sup>7</sup> Terzaghi

<sup>8</sup> Hottmann and Johnson

<sup>9</sup> Jorden and Shirley

در سال ۱۹۷۵ وایز و بایوت<sup>۱۰</sup> رابطه‌ای تجربی (معادله ۵) میان فشار سرباره، تنش مؤثر و فشار منفذی به همراه ضریبی به نام ضریب بایوت ارائه می‌دهند. در این رابطه ضریب بایوت نسبت تغییرات حجم سیال منفذی به تغییرات حجم کل سنگ بوده و در صورتیکه سیال برای خارج شدن از سنگ مانعی نداشته باشد، این مقدار برابر یک است. شایان ذکر است که، تئوری فوق فقط برای سنگ‌های همگن کاربرد دارد [۱۷].

$$PP = \frac{S_{over} - S_{eff}}{\beta} \quad (5)$$

در این معادله  $\beta$  ضریب بایوت است.

بیشترین تحقیقاتی که در طی دوره‌های اخیر بر روی پیش‌بینی فشار منفذی انجام شده است بر روی اطلاعات مربوط به نگاره سرعت موج‌برشی، مقاومت و فشاری صورت گرفته که این معادلات توسط ایتون<sup>۱۱</sup> (۱۹۷۵) و باورز<sup>۱۲</sup> (۱۹۹۵) ارائه شده است [۱۸-۲۰]. به‌طوری‌که در سال ۱۹۷۵، ایتون دو معادله تجربی برای پیش‌بینی فشار منفذی با استفاده از نگاره‌های موج تراکمی-فشاری، موج برشی و نگاره‌های مقاومت به‌صورت معادلات ۶ و ۷ ارائه داد [۲۰-۲۲].

$$PP = S_{over} - (S_{over} - S_{hyd}) \left( \frac{\Delta_{ts}}{\Delta_t} \right)^q \quad (6)$$

$$PP = S_{over} - (S_{over} - S_{hyd}) \left( \frac{GR_{ts}}{GR_t} \right)^q \quad (7)$$

در این معادله‌ها،  $S_{hyd}$  گرادیان فشار منفذی هیدرواستاتیک؛  $\Delta_{ts}$  نگاره صوتی اندازه‌گیری شده در شیل با نگاره موج تراکمی (ناحیه‌ی روند نرمال در نمودار چاه)؛  $\Delta_t$  نگاره صوتی اندازه‌گیری شده در شیل با نگاره موج تراکمی؛  $GR_{ts}$  نگاره مقاومت در پروفایل نرمال فشار منفذی و  $GR_t$  نگاره مقاومت هستند.

بر اساس نتایج ارائه شده، این معادلات تنها برای همان نواحی مورد مطالعه پاسخ منطقی دارد [۲۳، ۲۴]. برای برطرف کردن نقص مذکور، اخیراً به‌منظور پیش‌بینی پارامترهای کلیدی مربوط به صنعت نفت و گاز، محققان بسیاری به هوش مصنوعی گرایش پیدا کرده اند [۲۵].

در سال ۲۰۱۰، وانگ و همکاران<sup>۱۳</sup> به‌منظور پیش‌بینی فشار منفذی از سه روش روند خطی<sup>۱۴</sup>، روش فرمول فیلیپون اصلی<sup>۱۵</sup> و الگوریتم ژنتیک ترکیبی بدون نرخ جهش<sup>۱۶</sup> استفاده کردند [۲۶]. پس از بررسی این سه روش و روش‌های مرسوم، مشخص گردید که الگوریتم ژنتیک ترکیبی بدون نرخ جهش بهترین دقت عملکرد را دارد. در سال ۲۰۱۳، هو و همکاران<sup>۱۷</sup> بر اساس اطلاعات مربوط به پنج چاه از دو میدان متفاوت، از روش شبکه عصبی مصنوعی<sup>۱۸</sup> به پیش‌بینی فشار منفذی پرداختند. بیشترین درصد خطا برای این الگوریتم در حد ۷،۱۵٪ بود. این مطالعه نشان می‌دهد که دقت عملکرد الگوریتم مذکور قابل ملاحظه است [۱].

در سال ۲۰۱۴، عابدین<sup>۱۹</sup> باهدف یافتن بهترین روش به‌منظور پیش‌بینی فشار منفذی از الگوریتم شبکه عصبی مصنوعی استفاده کرد. دقت عملکرد این الگوریتم برای داده‌های استفاده‌شده در این مقاله ۵ درصد بوده که نشان از کارایی بالای این الگوریتم داشت [۲۷]. هاریس و همکاران<sup>۲۰</sup> در سال ۲۰۱۷، با استفاده از الگوریتم شبکه عصبی احتمالی به پیش‌بینی فشار منفذی پرداختند. ایشان بیان داشتند که نتایج حاصل از این الگوریتم دارای دقت عملکرد بالاتر از روابطی است که در آن‌ها از داده‌های لرزه‌ای استفاده می‌شود [۲۸].

<sup>10</sup> Biot and Willis

<sup>11</sup> Eaton

<sup>12</sup> Bowers

<sup>13</sup> Wang

<sup>14</sup> Trend line method

<sup>15</sup> The original Fillippone formula method

<sup>16</sup> Hybrid genetic algorithm without mutation rate

<sup>17</sup> Hu et al.

<sup>18</sup> Back Propagation Artificial Neural Network

<sup>19</sup> Abidin

<sup>20</sup> Haris et al.

در سال ۲۰۱۸، رشیدی و اسدی<sup>۲۱</sup> براساس سه چاه حفاری شده در سازند ماسه‌سنگی موجود در ایران که داده‌های آن شامل انرژی خاص مکانیکی<sup>۲۲</sup> و راندمان حفاری<sup>۲۳</sup> بود، با استفاده از مدل شبکه عصبی مصنوعی به پیش‌بینی فشار منفذی سازند پرداختند. نتایج مدل پیشنهادی هوش مصنوعی مورد تجزیه و تحلیل قرار گرفت و ثابت شد که شبکه‌های عصبی مصنوعی قادر به پیش‌بینی فشار منفذی هستند و این مدل هوشمند را می‌توان برای تجزیه و تحلیل داده‌ها برای ساخت مدل‌های پیش‌بینی پیش از عملیات حفاری استفاده کرد [۲۹].

کارماکار و مایاتی<sup>۲۴</sup> در سال ۲۰۱۹، بر اساس ۳۵۷ مجموعه داده‌ای و با استفاده از سه روش هوش مصنوعی بیضوی<sup>۲۵</sup> بهینه‌شده با مقیاس گرادینان مثبت<sup>۲۶</sup> و هیبرید مونت کارلو<sup>۲۷</sup> فشار منفذی را تخمین زدند. پس از بررسی‌های صورت گرفته مشخص گردید که ترکیب روش هوش مصنوعی بیضوی بهینه‌شده با مقیاس گرادینان مثبت دارای دقت عملکرد بهتری نسبت به الگوریتم هیبریدی دیگر است و مقدار خطای نسبی حدود ۰/۹۸ داشت [۳۰].

در سال ۲۰۲۰، یو و همکاران<sup>۲۸</sup> از چهار مدل هوش مصنوعی ماشین‌های تقویت گرادینان<sup>۲۹</sup>، دستگاه بردار پشتیبان<sup>۳۰</sup>، پرسپترون چندلایه<sup>۳۱</sup> و جنگل تصادفی<sup>۳۲</sup> به منظور پیش‌بینی فشار منفذی استفاده نمودند. نتایج تحقیقات آن‌ها نشان داد که جنگل تصادفی بهترین دقت عملکردی را دارا است [۲]. یک سال بعد یعنی در سال ۲۰۲۱، احمد عبدالله و همکاران<sup>۳۳</sup> از الگوریتم شبکه عصبی مصنوعی برای ۳۱۰۰ داده میدانی استفاده کرده‌اند که پس از ارائه نتایج مشخص گردید که میزان مربع ضریب همبستگی ۰/۹۸ و متوسط درصد مطلق خطای نسبی ۲ درصد است [۳۱].

بررسی ادبیات موجود نشان می‌دهد که بیشتر مطالعات مربوطه بر سازندهای تحت سلطه ماسه‌سنگ متمرکز شده‌اند، که بر اساس آن مدل‌ها برای توضیح رفتار شیل، یعنی همبستگی قوی بین فشار منفذی و تخلخل ساخته شده‌اند. با این حال، در یک مخزن کربناته، جایی که درجه ناهمگنی بسیار بالاتر است و سنگ مخزن به طور قابل توجهی سفت تر است، به طوری که فشار منفذی‌های بالا بدون ارتباط منطقی با تخلخل مواجه می‌شوند، چنین مدل‌های معمولی ممکن است در نهایت تخمین‌های با دقت پایینی تولید کنند. با این وجود، این مدل‌ها در حال حاضر برای پیش‌بینی فشار منفذی در صنعت نفت و گاز استفاده می‌شوند. بنابراین، پژوهش حاضر مدل‌سازی را بر اساس داده‌های سنگ‌های کربناته (ناهمگن) استوار می‌کند. با توجه به پیچیدگی ذاتی رفتار فشار منفذی و این واقعیت که این پیچیدگی در مخازن کربناته حتی بیشتر می‌شود، استفاده از تکنیک‌های هوش مصنوعی که به ویژه برای مدل‌سازی روابط غیرخطی پیچیده مفید بوده‌اند، از اهمیت بالایی برخوردار است. علاوه بر این، بررسی ادبیات نشان می‌دهد که هیچ تحقیق قبلی از الگوریتم جنگل تصادفی و الگوریتم رگرسیون بردار پشتیبان برای پیش‌بینی فشار منفذی استفاده نکرده است. این در حالی است که استفاده از این الگوریتم در سایر حوزه‌های صنعتی، قابلیت‌های بزرگ این الگوریتم را برای مدل‌سازی انجمن‌های غیرخطی پیچیده برجسته کرده است. از جمله مزیت‌های این الگوریتم‌ها، ارائه بهترین نتیجه بدون نیاز به دانش آماری، جدا کردن داده‌های غیرضروری، آماده‌سازی داده‌ها در زمان کوتاه و کاهش خطای نسبی با یافتن گره اصلی تصمیم گیر می‌باشد.

<sup>21</sup> Rashidi and Asadi

<sup>22</sup> Mechanical specific

<sup>23</sup> Drilling efficiency

<sup>24</sup> Karmakar and Maiti

<sup>25</sup> Bayesian neural networks

<sup>26</sup> Scaled Conjugate Gradient

<sup>27</sup> Hybrid Monte Carlo

<sup>28</sup> Yu et al.

<sup>29</sup> Gradient boosting machines

<sup>30</sup> Support vector machine

<sup>31</sup> Multilayer perceptron

<sup>32</sup> Random forest

<sup>33</sup> Abdelaal

## ۲- روش انجام کار

هوش مصنوعی در زبان کامپیوتری هوش ماشینی اطلاق می‌شود و در اکثر مقالات این ابزار به‌عنوان دانش شناخت و طراحی عامل‌های هوشمند شناخته می‌شود که در مقابل هوش طبیعی در انسان‌ها قرار دارد [۳۲، ۳۳]. به‌عبارت‌دیگر هوش مصنوعی به ابزاری گفته می‌شود که می‌تواند واکنش‌هایی مشابه رفتارهای هوشمند انسانی از جمله، درک شرایط پیچیده، شبیه‌سازی فرایندهای تفکری و شیوه‌های استدلالی انسانی و پاسخ موفق به آن‌ها، یادگیری و توانایی کسب دانش و استدلال برای حل مسائل را داشته باشند [۳۴]. الگوریتم‌های هوش مصنوعی امروزه به‌طور گسترده‌ای برای حل بسیاری از چالش‌های مهندسی مورد استفاده قرار گرفته است. ریشه‌ها و ایده‌های اصلی الگوریتم‌های هوش مصنوعی را باید در فلسفه، زبان‌شناسی، ریاضیات، روان‌شناسی، عصب‌شناسی، فیزیولوژی، تئوری کنترل، احتمالات و بهینه‌سازی جستجو کرد.

## ۲-۱- نمودار گردش کار

شکل ۱ شماتیکی از نمودار گردش کار است که خلاصه‌ای از مراحل مربوط به ساخت، ارزیابی و مقایسه روش‌های مصنوعی را به‌منظور پیش‌بینی فشار منفذی نشان می‌دهد. بر اساس این نمودار ابتدا داده‌ها را جمع‌آوری کرده، و پس از مرتب کردن داده‌ها، داده‌های پرت را با استفاده از روش *k-means clustering* حذف نموده‌ایم. تعداد داده‌های حذف شده برای این مقاله ۵۷ عدد می‌باشد که پس از بررسی روش تعیین می‌گردد. پس از آن برای هر متغیر مقدار حداقل ( $d_{min}^l$ ) و حداکثر ( $d_{max}^l$ ) را تعیین کرده و سپس فرایند نرمال‌سازی (معادله ۸) را انجام می‌دهیم. لازم به ذکر است داده‌های نرمال‌ایز شده، عددی بین ۰ و ۱ می‌باشند.

$$d_i^l = \left( \frac{d_i^l - d_{min}^l}{d_{max}^l - d_{min}^l} \right) * 2 - 1 \quad (8)$$

که در آن:  $d_i^l$  مقدار ویژگی ۱ برای داده  $i$ ام؛  $d_{min}^l$  کمترین مقدار ویژگی ۱ برای هر متغیر ورودی برای کل داده‌های مربوط به همان ویژگی؛  $d_{max}^l$  بیشترین مقدار ویژگی ۱ برای هر متغیر ورودی برای کل داده‌های مربوط به همان ویژگی است.

پس از نرمال‌سازی، داده‌ها به سه بخش آموزشی (۷۰٪)، تستی (۱۵٪) و اعتبارسنجی (۱۵٪) تقسیم‌بندی می‌شوند. نتایج حاصل شده به صورت بازه‌هایی بین ۰-۱ می‌باشد و در این مرحله می‌بایست مقدار واقعی هر پارامتر را به همان حالت اولیه خود در می‌آوریم (اعداد واقعی). در مرحله‌ی بعد، داده‌های حاصل، با استفاده از معیارهای آماری مدل‌های هوش مصنوعی مورد مقایسه قرار گرفته و بهترین الگوریتم مشخص می‌گردد. در پایان، الگوریتم مذکور را بر روی اطلاعات مربوط به یک چاه دیگر نیز تست کرده و نتیجه نهایی گزارش می‌شود.



شکل ۱: نمودار گردش کار برای پیش‌بینی نرخ فشار منفذی.

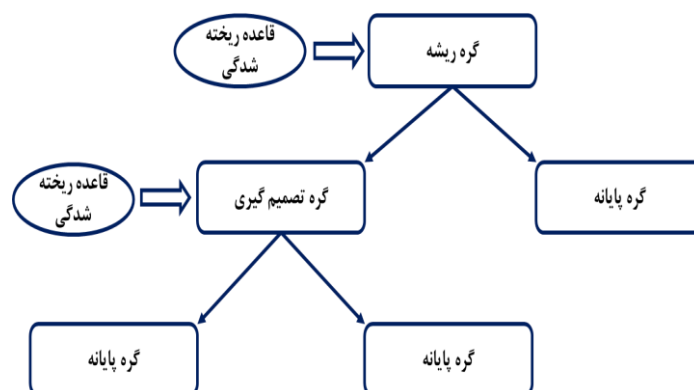
## ۲-۲- الگوریتم درخت تصمیم

یکی از روش‌ها یادگیری ماشین که به‌طور گسترده‌ای به‌منظور ارزیابی مجموعه داده‌ها مورداستفاده قرار می‌گیرد، درخت تصمیم است [۳۵، ۳۶]. به‌طوری‌که مجموعه‌ای از داده‌ها در یک ساختار سلسله‌مراتبی متشکل از گره‌ها و رشته‌ها، سازمان‌دهی شده، سپس توسط مجموعه‌ای از قوانین داده‌ها طبقه‌بندی و برای فرایند عددی (رگرسیون) آماده می‌شوند [۳۷]. به‌منظور ساخت درخت تصمیم برای برنامه‌های یادگیری ماشین ابتدا تمایزی بین متغیرهای ورودی (یا ویژگی) و متغیرهای هدف ایجاد کرده، سپس داده‌ها را بر اساس قوانین تعریف‌شده در گره‌های "کودک" تقسیم می‌کند. درنهایت با تفکیک بیشتر، هر گره کودک به‌عنوان گره والد عمل کرده که از آن گره‌های بیشتر و لایه‌های درخت تصمیم‌گیری ایجاد می‌شود [۳۸].

شکل ۲ یک درخت تصمیم‌گیری را نشان می‌دهد که انواع مختلف گره‌ها را باهم ترکیب می‌کند. نحوه‌ی کار این الگوریتم به این شکل است که از گره ریشه (لایه بالایی)، کل مجموعه داده را ابتدا به دو زیرمجموعه تقسیم کرده (گره‌های کودک در لایه دوم) سپس گره‌های کودک بیشتری به زیر گره‌های کوچک‌تر تقسیم می‌شوند و لایه‌های بیشتری از گره‌های کودک را تشکیل داده و درنهایت به یک‌لایه نهایی با گره‌های انتهایی (یا برگ) می‌رسند. تا زمانی که داده‌ها به‌صورت کامل توسط درخت تصمیم طبقه‌بندی نشوند، توسعه‌ی درخت، ادامه پیدا می‌کند. یکی از نکات مهم در مورد الگوریتم درخت تصمیم این است که هرچه لایه‌ها و گره‌های بیشتری در یک درخت تصمیم‌گیری وجود داشته باشد باعث پدیده بیش‌برازش<sup>۳۴</sup> گشته و دقت عملکرد کاهش می‌یابد [۳۹، ۴۰].

از مزایای مهم این الگوریتم خواندن و تفسیر ساده، آماده‌سازی آسان و جدا کردن داده‌هایی است که کمتر موردنیاز هستند. از دیگر مزایای درختان تصمیم‌گیر این است که خروجی آن‌ها بدون نیاز به دانش آماری قابل‌خواندن و تفسیر می‌باشد. همچنین، در مقایسه با سایر تکنیک‌های تصمیم‌گیری، برای آماده‌سازی داده‌ها، درختان تصمیم تلاش کمتری می‌کنند. با این حال، کاربران باید اطلاعات آماده‌ای برای ایجاد متغیرهای جدید، با قدرت پیش‌بینی متغیر هدف در اختیار داشته باشند. آن‌ها همچنین می‌توانند طبقه‌بندی داده‌ها را بدون محاسبات پیچیده ایجاد کنند. مزیت دیگر درخت تصمیم این است که پس از ایجاد متغیرها، داده‌هایی که کمتر موردنیاز هستند را جدا می‌کند.

<sup>34</sup> Overfitting



شكل ۲: شماتيك الءوریتم درءء تصمیم برای پیش بینی فشار منءذی.

در این مطالعه، یک مدل درء تصمیم گیری در نرم افزار پایتون، با استفاده از معیار gini جهت تعیین اهمیت هر ویژگی، کءذاری شده است. ءءول ۱، "بهترین" تقسیم کننده برای تصمیم گیری در مورد هر ویژگی و مقدار آستانه ای که در ایجاد هر تقسیم باید اعمال شود را نشان می دهد.

ءءول ۱: پارامترهای کنءرلی برای الءوریتم درءء تصمیم به منظور پیش بینی فشار منءذی.

مءءار	پارامترهای کنءرلی
۱۰۵	ءءاکءر عمق
Gini	معیار
Best	ءءاکننده
ءطای میانگین مربعات	ءابع هءء
۰،۰۱۱۴۵۷ ثانیه	مءال زمان پیش بینی

### ۳-۲- الءوریتم ماشین برءار پشتیبان

در سال ۱۹۹۵ الءوریتم ماشین برءار پشتیبان توسط كرت و وینیک بر اساس نظریه یادگیری آماری، ارائه شد. این الءوریتم یکی از الءوریتم های بسیار پرءارءر در زمینه های مختلف طبقه بندی و رءرسیون است [۴۱]. این الءوریتم ابتدا ءوابع هسته ای را برای نقشه برءاری برءارهای غیرءطی به ابعاد بالاتر مورد استفاده قرار می دهد. سپس، یک ابرهواپیما در فضای ویژگی ها ساخته شده و در نهایت ءاده ها را به ءو كلاس تقسیم بندی می كند. با ایجاد برءارهای پشتیبان، فاصله بین كلاس ها و ابرهواپیما برای زیرمجموعه آموزش به ءءاكثر مقدار خود می رسند [۴۲]. در این مقاله، یک مدل ماشین برءار پشتیبان با استفاده از رءرسیون، به منظور پیش بینی فشار منءذی ارائه شده است.

مدل ماشین برءار پشتیبان به منظور تعیین متغیرهای ورودی و متغیر خروجی مرتبط به ترتیب از پارامترهای  $X = R^h$  و  $y_i \in X = R$  استفاده می كند در جایی که  $i=1, 2, 3, \dots, h$  می باشد. پیش بینی های مربوط به این مدل هوش مصنوعی از یک ءابع رءرسیون  $y = f(x)$  به ءءست آمده و مقادیر هءء با استفاده از یک ءابع هءء یادگیری تقریب زءه می شود (معءءله ۹).

$$f(x, e) = hZ(x) + d \quad (9)$$

که در آن،  $f(x, e)$  پیش بینی ءابع هءء،  $Z(x)$  نقشه برءاری از ویژگی های فضای بالای  $x$ ،  $h \in R$  فاكءور وزن الءوریتم و  $d \in R$  ءء آستانه هر متغیر ورودی است.

ضرایب  $Z$  و  $d$  به شكل زیر و با كاهش عملكرد ریسك ءاصل می شود (معءءله ۱۰). همچنین به منظور تعریف ءابع هءء از معءءله ۱۱ استفاده می شود.



$$\frac{1}{2} \|Z\|^2 + C \frac{1}{l} \sum_{i=1}^n L_{\varepsilon}(y_i, f(x_i, Z)) \quad (10)$$

$$L_{\varepsilon}(y_i, f(x_i, Z)) = \begin{cases} 0, & y_i - f(x_i, Z) \leq \varepsilon \\ |y_i - f(x_i, Z)| - \varepsilon, & \text{otherwise} \end{cases} \quad (11)$$

در این معادله‌ها،  $\|Z\|^2$  صاف بودن عملکرد،  $C$  معیاری میان میزان مبادله بین مسطح بودن مدل و خطای تجربی و  $\frac{1}{l} \sum_{i=1}^n L_{\varepsilon}(y_i, f(x_i))$  نشانگر تابع هدف است.

یکی از راه‌های تبدیل ضرایب  $Z$  و  $d$  به تابع هدف اصلی، استفاده از تبدیل تابع خطی منظم و تعریف دو ثابت  $\varphi^*$  و  $\varphi$  به شرح زیر است:

$$\text{Min } \frac{1}{2} Z^2 + C \frac{1}{l} \sum_{i=1}^l (\varphi_i - \varphi_i^*) \quad (12)$$

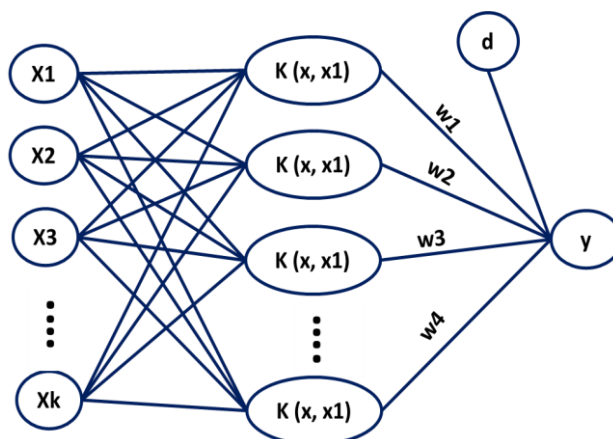
$$\text{Subjected to } \begin{cases} y_i - f(x_i, Z) - b \leq \varepsilon + \varphi_i \\ f(x_i, Z) + d - y_i \leq \varepsilon + \varphi_i^* \\ \varphi_i, \varphi_i^* > 0 \end{cases} \quad (13)$$

در نهایت این معادله را می‌توان با استفاده از ضرب لاگرانژی حل کرد به طوری که پاسخ با معادله ۱۴ بیان می‌شود [۴۳، ۴۴].

$$f(x, \rho_i, \rho_i^*) \sum_{i=1}^N (\rho_i - \rho_i^*) (\forall(x_i), \forall(x)) + d \quad (14)$$

به منظور غلبه بر پیچیدگی‌های محاسبات ریاضی در فضایی با ابعاد بالا، باید تابع هسته‌ای مناسب تعریف کرد. این هسته‌ها می‌تواند علاوه بر حل پیچیدگی‌های محاسباتی، سرعت و پردازش عملکرد را نیز کنترل کند. به طور کلی چهار تابع هسته‌ای برای این الگوریتم وجود دارد که شامل چند جمله‌ای<sup>۳۵</sup>، سیگموئید<sup>۳۶</sup>، تابع پایه شعاعی<sup>۳۷</sup> و خطی<sup>۳۸</sup> هستند [۴۵].

از میان هسته‌های ذکر شده، از هسته‌ی تابع پایه شعاعی در این مطالعه استفاده شده است (شکل ۳). از ویژگی‌های هسته مذکور این است که می‌تواند از نویزهای به وجود آمده در داده‌های مربوط به آموزش جلوگیری به عمل آورد [۴۶]. جدول ۲، پارامترهای کنترلی مربوط به این الگوریتم را نشان می‌دهد.



شکل ۳: شماتیک الگوریتم ماشین بردار پشتیبان برای پیش‌بینی فشار منفذی.

<sup>35</sup> Polynomial

<sup>36</sup> Sigmoid

<sup>37</sup> Radial basis function

<sup>38</sup> Linear

جدول ۲: پارامترهای کنترلی برای الگوریتم ماشین بردار پشتیبان به منظور پیش بینی فشار منفذی.

پارامترهای کنترلی	مقدار
عملکرد هسته	RBF
محدوده $\epsilon$	۰/۲
محدوده C	۱۰۵۰۰۰
اعتبار سنجی متقابل	اعمال نشده است
$\gamma$ محدوده (RBF)	۰/۰۴۵

#### ۴-۲- الگوریتم جنگل های تصادفی

یکی از الگوریتم‌هایی که شباهت زیادی به الگوریتم درخت تصمیم دارد، الگوریتم جنگل‌های تصادفی است. به گونه‌ای که این الگوریتم را می‌توان، توسعه یافته الگوریتم درخت تصمیم نامید. این الگوریتم، چندین درخت تصمیم را برای ارزیابی ایجاد و توسعه می‌دهد. الگوریتم مذکور، یک الگوریتم تحت نظارت است که اساس آن بر طبقه‌بندی، رگرسیون و آموزش مبتنی بر داده‌های متغیرهای ورودی بنا شده است [۴۷، ۴۸]. طریقه ساخت درختان تصمیم‌گیری متعدد به صورت موازی و رو به روی هم است و به منظور ارتباط بین آن‌ها از لایه‌ها و گره‌های نسبتاً کمی استفاده می‌شود. این الگوریتم نیز مشابه الگوریتم درخت تصمیم نسبت به افزایش گره و لایه حساس است و افزایش آن سبب پدیده بیش برآزش در دقت عملکرد الگوریتم می‌شود [۴۹].

به منظور آموزش الگوریتم جنگل تصادفی، ابتدا بخشی از یک مجموعه داده باید انتخاب و بوت استرپ<sup>۳۹</sup> گردد ( به صورت تصادفی). سپس می‌توان با استفاده از این مجموعه، برای دسته‌بندی داده‌های طبقه‌بندی نشده و یا درخت رگرسیون استفاده کرد. یکی از ویژگی‌های بارز این الگوریتم این است که همه‌ی متغیرهای ورودی را به عنوان کاندیدا انتخاب نمی‌کند. این الگوریتم، تعداد کمی از متغیرهای ورودی‌ها را به طور تصادفی پذیرفته و سپس برای عملیات آموزش استفاده می‌کند. نکته حائز اهمیت این است که اگر تعداد درختان به اندازه کافی نباشند، این الگوریتم چنین قابلیت‌هایی دارد تا تعدادی درخت تصمیم‌گیری به صورت تکرار بسازد تا پیش بینی متغیر وابسته (تابع هدف) به خوبی صورت پذیرد. تابع پیش بینی الگوریتم جنگل تصادفی به صورت زیر بیان می‌گردد:

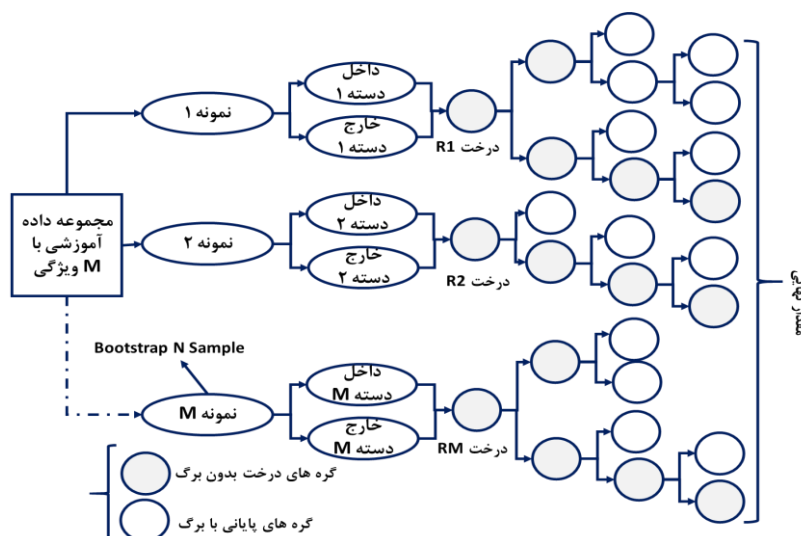
$$\hat{f}_{RF}^M(x) = \frac{1}{M} \sum_{k=1}^M R_k(x) \quad (15)$$

که در آن  $\hat{f}_{RF}^M(x)$  تابع پیش بینی الگوریتم جنگل تصادفی،  $M$  تعداد درختان مستقل،  $x$  تعداد متغیرهای ورودی و  $R_k(x)$  پیش بینی رگرسیون یک درخت برای  $i$  امین داده است.

به منظور برآورد خطای درختان خارج از دسته باید از تک تک درختان به صورت مجزا رگرسیون گرفته سپس میانگین نتایج گزارش گردد. درختان خارج از دسته آن قسمت از داده‌هایی هستند که در فرایند آموزش انتخاب نشده و پس از آموزش در فرایند دسته بندی مشتق می‌شوند. این الگوریتم، به منظور بررسی خطای درختان خارج از دسته، ابتدا از صحت پیش بینی نتایج حاصل از داده‌های خارج از دسته مطلع می‌شود (این مجموعه را به صورت یک زیرمجموعه مستقل ارائه داده و با نتایج درون دسته مقایسه و صحت سنجی می‌کند). سپس، اهمیت نسبی هر یک از متغیرهای ورودی را نسبت به متغیر هدف بر اساس جمع بندی نتایج ارزیابی می‌کند. این امر به حذف و کاهش متغیرهای ورودی به منظور کاهش ابعاد مدل کمک می‌کند تا کارایی مدل، سرعت پردازش و همچنین دقت عملکرد نتایج حاصله افزایش یابد [۵۰]. از جمله راه‌هایی که برای دستیابی به این هدف وجود دارد این است که با تغییر دو متغیر ورودی و ثابت نگه داشتن بقیه ی متغیرهای باقیمانده، ریشه میانگین خطای مربع<sup>۴۰</sup> را ثبت و اهمیت نسبی هر متغیر ورودی برای متغیر وابسته را مشخص کرد [۴۹]. شکل ۴ شماتیک الگوریتم جنگل تصادفی را نشان می‌دهد.

<sup>39</sup> Bootstrapped

<sup>40</sup> Root mean square error



شکل ۴: شمانیک روش الگوریتم جنگل تصادفی برای درخت‌های درون دسته و خارج از دسته.

در این مطالعه برای پیش‌بینی فشار منفذی برای چاه‌های یکی از میدان‌ها جنوب ایران از محیط Scikit Learn الگوریتم جنگل تصادفی استفاده شده که پارامترهای کنترلی مربوط به این الگوریتم در جدول ۳ آمده است.

جدول ۳: پارامترهای کنترلی برای الگوریتم جنگل تصادفی به منظور پیش‌بینی فشار منفذی.

مقدار	پارامترهای کنترلی
۱۰۵۰	حداکثر عمق
.	حالت تصادفی
۱۰۵۰	تعداد درختان تصمیم‌گیری
خطای میانگین مربعات	تابع هدف
۵/۴۲	مثال زمان پیش‌بینی

## ۵-۲- الگوریتم شبکه عصبی مصنوعی

یکی از پرکاربردترین الگوریتم‌ها در علوم مختلف که به بررسی و حل مشکلات پیچیده ریاضی کمک شایانی می‌کند الگوریتم شبکه عصبی مصنوعی است [۵۱]. این الگوریتم یکی از الگوریتم‌های جعبه سیاه<sup>۴۱</sup> بوده به گونه‌ای که درون این الگوریتم‌ها لایه‌های پنهان قرار می‌گیرند تا محاسبات رگرسیون‌های پیچیده را با دقت بیشتری انجام دهد. در این مطالعه از الگوریتم شبکه عصبی مصنوعی استفاده شده که دارای یک لایه پنهان می‌باشد.

انتقال اطلاعات از سلول‌های عصبی یک لایه به لایه‌ی بعدی در الگوریتم شبکه عصبی مصنوعی بر اساس دو پارامتر وزن و بایاس صورت می‌گیرد که با تنظیم آن‌ها می‌توان به عملکرد مطلوب رسید. اطلاعات وارد شده به الگوریتم شبکه عصبی مصنوعی در لایه پنهان پردازش شده و با استفاده از یک تابع فعال‌ساز تنظیم می‌شوند. تابع فعال‌ساز برای الگوریتم شبکه عصبی مصنوعی توسط معادله ۱۶ معرفی شده است:

$$f(x) = f\left(\sum_{i=1} W_{ij}x_i + b_j\right) \quad (16)$$

که در آن،  $f(x)$  تابع فعال‌ساز،  $b$  لایه بایاس،  $x_i$  متغیر ورودی  $i^{\text{th}}$ ،  $W_i$  وزن ارتباط بین ورودی‌ها  $i^{\text{th}}$  و نرون  $i^{\text{th}}$  است.

<sup>41</sup> Black box

شبکه‌ی عصبی معمولاً با الگوریتم انتشار مجدد<sup>۴۲</sup> تنظیم وزن‌ها و بایاس‌ها را در لایه پنهان بهبود بخشیده و این مهم را با به حداقل رساندن میانگین ریشه خطای مربع بین مقادیر پیش‌بینی و واقعی برای همه داده‌های آموزشی طبق معادله ۱۷ انجام می‌دهد.

$$Error_{MSE} = \frac{1}{S} \sum_i^m (\hat{\beta}_i - \beta_i)^2 \quad (17)$$

که در آن،  $Error_{MSE}$  خطای میانگین ریشه خطای مربع،  $S$  تعداد داده‌های موجود برای هر پارامتر،  $\hat{\beta}_i$  مقدار واقعی برای هر متغیر ورودی،  $\beta_i$  مقدار پیش‌بینی برای هر متغیر ورودی می‌باشد.

در این مطالعه از بهینه‌ساز RMSprop استفاده شده است. نرخ یادگیری اولیه مربوط به RMSprop که یک الگوریتم مبتنی بر گرادیان است در معادلات ۱۸ و ۱۹ نشان داده شده‌اند [۵۲].

$$\gamma[gr^2]_n = 0.9E[gr^2]_{n-1} + 0.1gr_n^2 \quad (18)$$

$$\tau_{n+1} = \tau_n - \frac{\delta}{\sqrt{\gamma[gr^2]_n + \epsilon}} gr_n \quad (19)$$

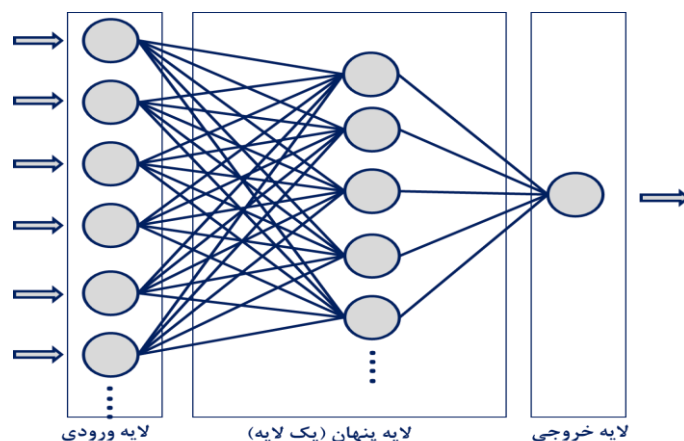
که در آن،  $\gamma[gr^2]_n$  گرادیان میانگین برای  $n$  تکرار،  $gr_n$  تابع گرادیان برای  $n$  تکرار،  $\tau_n$  تابع هدف برای  $n$  تکرار،  $\delta$  میزان یادگیری برای الگوریتم می‌باشد.

ساختار شبکه عصبی مصنوعی شامل یک تک لایه مخفی است که همان تعداد نورون که به عنوان متغیرهای ورودی وارد شبکه عصبی می‌شود را پردازش و ۶۰۰ وزن در یک لایه مخفی و با یک لایه خروجی با ۱ نورون برای پیش‌بینی متغیر وابسته (هدف) (شکل ۵). پیکربندی و اجرای این مدل در Keras مربوط به یک بسته یادگیری عمیق کدگذاری شده در پایتون صورت گرفته که بر روی محیط یادگیری ماشین TensorFlow اجرا می‌شود. جدول ۴ پارامترهای کنترلی مربوط به این الگوریتم گزارش شده است.

جدول ۴: پارامترهای کنترلی برای الگوریتم ANN به منظور پیش‌بینی فشار منفذی.

پارامترهای کنترلی	مقدار
تعداد لایه‌های پنهان	۱
تعداد وزن‌ها در لایه پنهان	۶۰۰
تابع فعال‌سازی از ورودی به لایه مخفی استفاده می‌کند	SELU (واحد خطی مقیاس نمایی)
تابع فعال‌سازی برای مخفی کردن لایه خروجی استفاده می‌شود	SELU
عملکرد هدف برای زیرمجموعه آموزشی به حداقل رسیده است	MSE
الگوریتم بهینه‌سازی	RMSprop
حداقل دلتا	٪۲۰
تعداد تکرارها	۴۰۲
میزان یادگیری	۱۵۰

<sup>42</sup> Backpropagation algorithm



شکل ۵: شماتیک الگوریتم شبکه عصبی مصنوعی برای پیش‌بینی فشار منفذی.

## ۲-۶- انتخاب ویژگی و روش‌های مربوط به انتخاب ویژگی

به‌منظور یافتن بهترین و تأثیرگذارترین متغیرهای ورودی، می‌توان دقت عملکرد پیش‌بینی فشار منفذی را با استفاده از الگوریتم‌های جدید هوش مصنوعی افزایش و زمان اجرای الگوریتم‌ها را نیز به‌هنگام تجزیه و تحلیل کاهش داد [۵۳]. یکی از دلایل ارائه روش انتخاب ویژگی، وجود تعداد بسیار زیادی از متغیرهای ورودی بالقوه بوده و این متغیر بر سرعت و دقت الگوریتم‌های هوش مصنوعی تأثیر بسیار زیادی می‌گذارد [54]. جهت بهبود عملکرد مدل‌های هوش مصنوعی، تعداد ترکیب متغیرها تابعی از تعداد متغیر ورودی ( $X$ ) بوده و از قانون  $2^x$  تبعیت می‌کند [55, 56]. به عبارتی اگر به‌منظور پیش‌بینی فشار منفذی از ۳ متغیر ورودی استفاده شود، ۲۰۴۸ ترکیب از آن ویژگی‌ها ایجاد می‌شود که این تعداد ترکیب متغیرها سبب کاهش سرعت عملکرد هوش مصنوعی می‌گردد. روش‌های انتخاب ویژگی، شامل سه روش فیلترینگ<sup>۴۳</sup>، بسته‌بندی<sup>۴۴</sup> و جاسازی<sup>۴۵</sup> بوده که بر اساس نتایج محققان، روش بسته‌بندی جز دقیق‌ترین روش‌ها محسوب می‌شود [۵۷، ۵۸]. در این روش، از هیبرید الگوریتم ژنتیک و پرسپترون ساده چندلایه به‌منظور پیش‌بینی فشار منفذی باهدف کاهش تابع هزینه خطای میانگین مربع استفاده می‌شود [۲۲، ۵۹، ۶۰].

## ۳- جمع‌آوری داده‌ها، بررسی انتخاب ویژگی، توصیف داده‌ها

### ۳-۱- جمع‌آوری داده‌ها

در این مقاله از داده‌های مربوط به سه چاه #A، #B و #C واقع در یکی از میدان‌ها نفتی جنوب غرب ایران استفاده شده است. تعداد داده‌های مورد استفاده در چاه #A ۹۸۸ داده در فاصله عمق بین ۳۲۵۷ تا ۳۴۵۴ متر، تعداد داده‌های مورد استفاده در چاه #B ۹۰۵ داده در فاصله عمق بین ۳۱۹۴ تا ۳۳۷۵ متر و تعداد داده‌های مورد استفاده در چاه #C ۹۳۴ داده در فاصله عمق بین ۳۲۰۴ تا ۳۳۹۰ متر می‌باشند. لازم به ذکر است که فاصله ثبت داده‌ها برای هر سه چاه #A، #B و #C ۰/۲ متر است. در ابتدا به‌منظور پیش‌بینی فشار سازندی، ۱۲ متغیر شامل نگاره القایی سطحی<sup>۴۶</sup> (LLS)، تصحیح اشعه گاما<sup>۴۷</sup> (CGR)، سرعت موج برشی<sup>۴۸</sup> ( $V_s$ )، اشعه گامای طیفی اصلاح‌نشده<sup>۴۹</sup> (SGR)، زمان انتقال صوتی فشرده‌سازی<sup>۵۰</sup> (DT)، تخلخل نوترونی<sup>۵۱</sup> (NPHI)، ضریب جذب فوتوالکتریک<sup>۵۲</sup> (PEF)، مقاومت

<sup>43</sup> Filtering

<sup>44</sup> Wrapping

<sup>45</sup> Embedded

<sup>46</sup> Laterolog shallow

<sup>47</sup> Corrected gamma ray

<sup>48</sup> Shear-wave velocity

<sup>49</sup> Uncorrected spectral gamma-ray

<sup>50</sup> Compression sonic transit time

<sup>51</sup> Neutron porosity

<sup>52</sup> The photoelectric absorption factors

عمیق<sup>۵۳</sup> (ILD)، تراکم سنگ<sup>۵۴</sup> (RHOB)، نگاره قطرسنج<sup>۵۵</sup> (CALL)، اندازه چاه<sup>۵۶</sup> (HS) و سرعت موج فشاری<sup>۵۷</sup> (Vp) توسط اداره پتروفیزیک شرکت ملی نفت فراهم شده است.

### ۲-۳- انتخاب ویژگی برای داده‌های ورودی

پس از تعیین الگوریتم MLP-GA به منظور استفاده در انتخاب ویژگی، ابتدا کل داده‌های مربوط به چاه‌های #A و #B (۱۸۹۳ داده ست) را به دلیل پراکندگی و وسعت داده‌های بیشتر به الگوریتم MLP-GA ارائه داده سپس این الگوریتم با توجه به دسته‌بندی‌هایی که صورت می‌دهد و همچنین معیاری که در نظر دارد (کمترین مقدار RMSE) به کاهش ویژگی‌ها کمک کرده تا بهترین ترکیب از مؤثرترین پارامترها را از نظر دقت عملکرد ارائه دهد. به عنوان مثال برای دسته‌های یکی، تک تک پارامترهای ورودی (H1, H2, ...) را به صورت جداگانه به عنوان ورودی به الگوریتم MLP-GA داده و خروجی گرفته، سپس با استفاده از معیار RMSE نتایج را با هم مقایسه کرده که بهترین نتیجه مربوط به H10 بوده و به همین ترتیب برای دسته‌های دیگر نیز به همین گونه عمل می‌کند. جدول ۵ نشانگرها را برای داده‌های ورودی نشان می‌دهد. نتایج موجود در جدول ۶ بیان‌کننده این واقعیت است که از ۱۲ ترکیب متغیر ورودی، ۹ متغیر ورودی LLS، CGR، Vs، Vp، SGR، NPHI، PEF، ILD و RHOB به عنوان بهترین ترکیب از نظر دقت عملکرد مطرح می‌باشند.

جدول ۵: نشانگرهای هر ویژگی به منظور پیش‌بینی فشار منفذی.

ویژگی	نماد	ویژگی	نماد
NPHI	H1	CALL	H7
HS	H2	SGR	H8
ILD	H3	PEF	H9
RHOB	H4	Vs	H10
CGR	H5	DT	H11
LLS	H6	Vp	H12

جدول ۶: نتایج ارائه‌شده از روش انتخاب ویژگی با استفاده از داده‌های مربوط به دو چاه #A و #B (۱۸۹۳ داده ست).

تعداد ویژگی	ویژگی‌های ورودی	مقدار RMSE
1	H10	122.1249
2	H10, H12	95.3524
3	H4, H12, H10	81.5487
4	H12, H3, H10, H4	72.3655
5	H6, H4, H10, H12, H3	65.1785
6	H8, H3, H6, H4, H10, H12	60.3248
7	H5, H10, H8, H12, H4, H3, H6	56.6555
8	H9, H8, H12, H6, H3, H4, H10, H5	53.3476
9	<b>H1, H6, H3, H8, H4, H10, H12, H5, H9</b> (بهترین ترکیب)	<b>51.9215</b>
10	H11, H9, H4, H10, H12, H5, H8, H3, H6, H1	53.7845
11	H2, H4, H1, H5, H10, H8, H9, H6, H11, H12, H3	56.0024
12	H7, H11, H12, H1, H3, H2, H8, H9, H5, H10, H4, H6	58.6315

### ۳-۳- توصیف داده‌ها

پس از بررسی انتخاب ویژگی برای متغیرهای ورودی و تعیین بهترین انتخاب برای ترکیب متغیرهای ورودی (۹ متغیر ورودی)، از اطلاعات مربوط به دو چاه #A و #B برای ساخت الگوریتم‌های هوش مصنوعی استفاده می‌شود. جدول ۷ پارامترهای آماری مربوط به

<sup>53</sup> Deep resistivity

<sup>54</sup> Bulk density

<sup>55</sup> Caliper

<sup>56</sup> Hole size

<sup>57</sup> Compressional-wave velocity

متغیرهای ورودی را پس از انتخاب ویژگی به ترتیب برای مجموع کل داده‌ها (سه چاه) ارائه داده است. به منظور اعتبار سنجی داده‌های مربوط به فشار منفذی از ابزار آزمایش مجدد سازند<sup>۵۸</sup> استفاده شده است.

جدول ۷. اطلاعات آماری مربوط به متغیرهای ورودی پس از بررسی روش انتخاب ویژگی برای کل داده‌های سه چاه #A, #B and #C (۲۲۲۷ مجموعه داده‌ای).

متغیرها	تصحیح اشعه گاما	تخلخل نوترونی	سرعت موج فشاری	ضریب جذب فوتوالکتریک	مقاومت عمیق	سرعت موج برشی	نگاره القایی سطحی	اشعه گامای طیفی اصلاح‌نشده	تراکم سنگ	فشار منفذی	چاه	نماد واحد
CGR	NPPI	Vp	PEF	ILD	Vs	LLS	SGR	RHOB	PP	فشار منفذی	چاه	نماد واحد
GAPI	PU	km/s	Barn/cm3	mmho/m	km/s	mmho/m	GAPI	g/cm3	Psi	فشار منفذی	چاه	نماد واحد
23.45	13.03	53.64	3.03	1226.05	246.57	209.35	45.00	2.98	5718.30	فشار منفذی	چاه	میانگین
19.51	5.50	2.96	1.48	4421.68	241.40	1631.94	20.92	0.54	401.51	فشار منفذی	چاه	انحراف
380.69	30.22	8.75	2.20	19544359.31	58251.17	2662282.39	437.54	0.29	161151.80	فشار منفذی	چاه	اطلاعات مربوط به معیار واریانس
1.06	-1.55	45.72	-0.45	0.42	57.55	0.48	12.21	1.20	4592.54	فشار منفذی	چاه	تمام چاه‌های مینیمم
124.27	46.67	82.91	6.33	20012.34	738.98	20003.12	146.30	3.93	6690.12	فشار منفذی	چاه	#A, #B و ماکزیمم
2.05	0.92	2.09	-0.42	3.85	0.92	11.24	1.45	0.21	-0.07	فشار منفذی	چاه	#C چولگی
4.78	4.84	12.90	-0.54	13.17	-1.09	130.26	2.94	-1.12	-0.53	فشار منفذی	چاه	چسبندگی

#### ۴- بحث و نتایج

به منظور بررسی و مقایسه‌ی الگوریتم‌های استفاده‌شده در این تحقیق به منظور پیش‌بینی فشار منفذی از پارامترهای آماری خطای نسبی، متوسط خطای نسبی، متوسط خطای نسبی مطلق، خطای میانگین مربع، ضریب تعیین، خطای میانگین مربع خطا و انحراف استاندارد استفاده‌شده است (معادلات ۲۰-۲۶).

. خطای نسبی (RE):

$$RE_i = \frac{PP_{(Measured)} - PP_{(Predicted)}}{PP_{(Measured)}} \times 100 \quad (20)$$

متوسط خطای نسبی (ARE):

$$ARE = \frac{\sum_{i=1}^n RE_i}{n} \quad (21)$$

متوسط خطای نسبی مطلق (AARE):

$$AARE = \frac{\sum_{i=1}^n |RE_i|}{n} \quad (22)$$

ضریب تعیین ( $R^2$ ):

$$R^2 = 1 - \frac{\sum_{i=1}^N (PP_{Predicted_i} - PP_{Measured_i})^2}{\sum_{i=1}^N (PP_{Predicted_i} - \frac{\sum_{i=1}^N PP_{Measured_i}}{n})^2} \quad (23)$$

خطای میانگین مربع (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (PP_{Measured_i} - PP_{Predicted_i})^2 \quad (24)$$

خطای میانگین مربع خطا (RMSE):

$$RMSE = \sqrt{MSE} \quad (25)$$

<sup>58</sup> Repeat formation tester tool

انحراف استاندارد (STD):

$$STD = \sqrt{\frac{\sum_{i=1}^n (D_i - Dimean)^2}{n-1}}$$

$$Dimean = \frac{1}{n} \sum_{i=1}^n (PP_{Measured_i} - PP_{Predicted_i}) \quad (26)$$

به منظور توسعه چهار الگوریتم جنگل تصادفی، الگوریتم رگرسیون بردار پشتیبان، الگوریتم شبکه عصبی مصنوعی و الگوریتم درخت تصمیم از ۱۸۹۳ داده‌ی جمع‌آوری شده از چاه #A و #B استفاده شده است که ۷۰٪ از داده‌ها برای آموزش، ۱۵٪ داده برای تست و ۱۵٪ داده‌ها جهت اعتبارسنجی مورد استفاده قرار گرفته است. نتایج مربوط به مقایسه الگوریتم‌ها در جدول ۸ ارائه شده است.

جدول ۸: دقت عملکرد پیش‌بینی فشار منفذی برای داده‌های آموزشی، تستی، اعتبارسنجی و کل داده‌ها با استفاده از داده‌های جمع‌آوری شده از چاه‌های #A و #B.

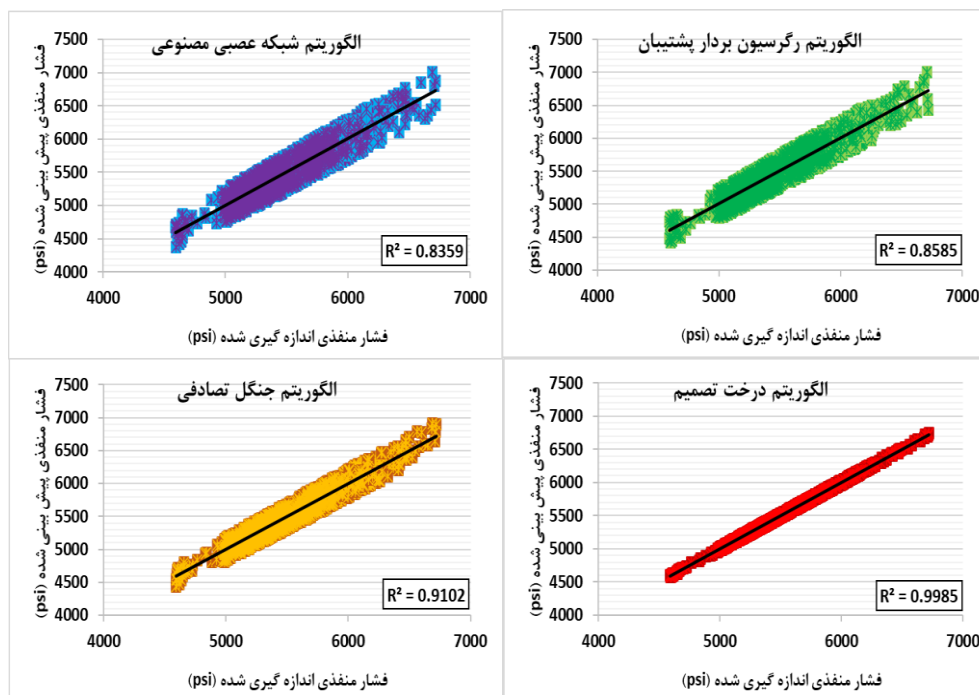
R <sup>2</sup>	RMSE	MSE	STD	AARE	ARE	مدل‌ها	پردازش
-	(Psi)	(Psi)	(Psi)	(%)	(%)	واحدها	واحدها
0.8069	143.4727	2.058E+04	143.541	1.908	-0.057	الگوریتم شبکه عصبی مصنوعی	مجموعه داده‌های آموزش: ۷۰٪ (۱۳۲۵) داده ست از کل داده‌های
0.8312	133.0981	1.772E+04	133.042	1.801	-0.096	الگوریتم رگرسیون بردار پشتیبان	
0.8951	101.3535	1.027E+04	101.311	1.351	0.045	الگوریتم جنگل تصادفی	
0.9980	14.3365	2.055E+02	14.336	0.191	-0.002	الگوریتم درخت تصمیم	
0.7167	164.3352	2.701E+04	164.232	2.870	0.115	الگوریتم شبکه عصبی مصنوعی	مجموعه داده‌های تستی: ۱۵٪ (۲۸۴) داده ست از کل داده‌های جمع‌آوری شده
0.7747	149.4746	2.234E+04	149.232	2.634	-0.161	الگوریتم رگرسیون بردار پشتیبان	
0.8482	114.7117	1.316E+04	114.449	1.974	-0.153	الگوریتم جنگل تصادفی	
0.9982	16.4298	2.699E+02	16.428	0.282	0.007	الگوریتم درخت تصمیم	
0.7535	171.5848	2.944E+04	171.530	3.058	0.086	الگوریتم شبکه عصبی مصنوعی	مجموعه داده‌های اعتبارسنجی: ۱۵٪ (۲۸۴) داده ست از کل داده‌های جمع‌آوری شده
0.8016	148.1807	2.196E+04	148.133	2.647	-0.082	الگوریتم رگرسیون بردار پشتیبان	
0.8921	116.2371	1.351E+04	116.222	2.037	-0.047	الگوریتم جنگل تصادفی	
0.9954	16.5256	2.731E+02	16.271	0.293	0.060	الگوریتم درخت تصمیم	
0.8359	145.3917	2.114E+04	145.611	2.039	-0.041	الگوریتم شبکه عصبی مصنوعی	مجموع کل داده‌ها: ۱۸۹۳ داده ست
0.8585	133.0242	1.770E+04	132.973	1.886	-0.068	الگوریتم رگرسیون بردار پشتیبان	
0.9102	101.6480	1.033E+04	101.657	1.420	0.032	الگوریتم جنگل تصادفی	
0.9985	14.4669	2.093E+02	14.470	0.202	-0.001	الگوریتم درخت تصمیم	

بر اساس نتایج حاصل در جدول ۸ مشخص می‌گردد که دقت عملکرد الگوریتم درخت تصمیم از دیگر الگوریتم‌ها بسیار بالاتر است. این عملکرد ناشی از ویژگی بارز الگوریتم درخت تصمیم در ارائه بهترین نتیجه بدون نیاز به دانش آماری و جدا کردن داده‌های غیرضروری می‌باشد.

یکی از نمودارهایی که به صورت بصری می‌تواند مقایسه‌ی بسیار خوبی برای الگوریتم‌ها ارائه دهد استفاده از نمودارهای متقاطع و همچنان استفاده از ضریب تعیین هست. همان‌گونه که در شکل ۶ و همچنین نتایج جدول ۸ مشاهده می‌شود میزان دقت عملکرد الگوریتم‌های هوش مصنوعی برای پیش‌بینی فشار منفذی، به صورت زیر است:

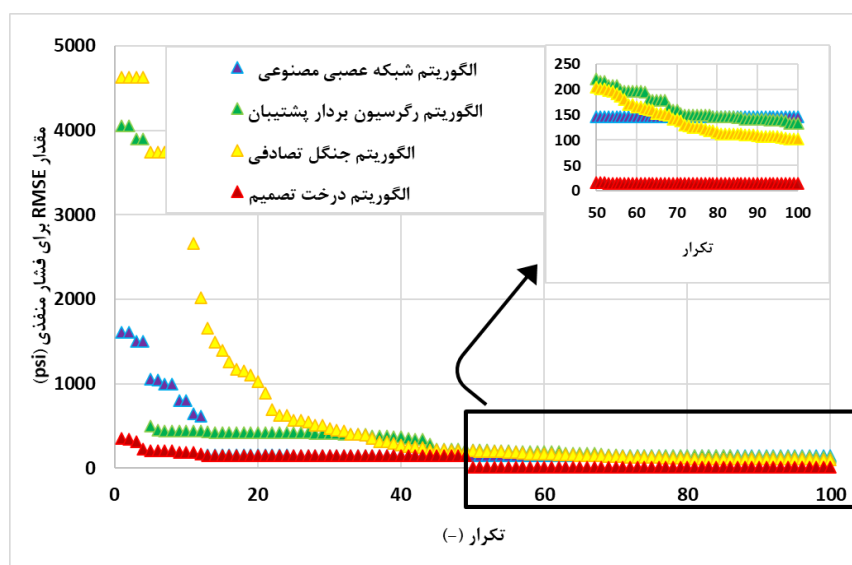
الگوریتم درخت تصمیم < الگوریتم جنگل تصادفی < الگوریتم رگرسیون بردار پشتیبان < الگوریتم شبکه عصبی مصنوعی





شکل ۶: نمودار متقاطع (cross plot) برای پیش‌بینی فشار منغذی برای چهار الگوریتم هوش مصنوعی به صورت مجزا برای کل داده‌های جمع‌آوری شده مربوط به چاه‌های A و B (# ۱۸۹۳ داده ست).

شکل ۷ مقادیر RMSE به ازای ۱۰۰ تکرار را برای چهار الگوریتم هوش مصنوعی نشان می‌دهد. همان‌گونه که در این شکل مشاهده می‌شود الگوریتم درخت تصمیم از ابتدا همگرا بوده و دقت عملکرد بالایی داشته (مقدار RMSE کم) و با همان روال کاهشی تا رسیدن به تکرار ۱۰۰ پیش می‌رود. الگوریتم جنگل تصادفی، الگوریتم رگرسیون بردار پشتیبان و الگوریتم شبکه عصبی مصنوعی ابتدا دارای RMSE بالا (دقت پیش‌بینی کم) سپس به سرعت این مقادیر همگرا شده و دقت عملکرد الگوریتم‌ها افزایش می‌یابد.



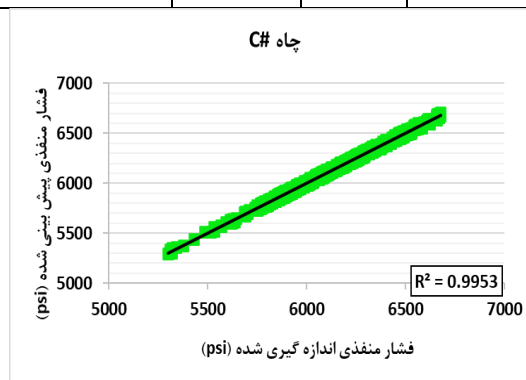
شکل ۷: مقایسه مقدار RMSE در مقابل تعداد تکرار برای چهار الگوریتم هوش مصنوعی به منظور پیش‌بینی فشار منغذی.

## ۵- تعمیم الگوریتم درخت تصمیم به منظور پیش‌بینی فشار منفذی برای چاه‌های دیگر میدان

به منظور تعمیم الگوریتم درخت تصمیم آموزش دیده با استفاده از اطلاعات مربوط به دو چاه #A و #B، از اطلاعات مربوط به چاه #C (۹۳۴ داده ست) استفاده می‌گردد. جدول ۹ و شکل ۸ نتایج مربوط به این الگوریتم را برای پیش‌بینی فشار منفذی نشان می‌دهد. بر اساس اطلاعات گزارش شده این الگوریتم قابلیت پیش‌بینی فشار منفذی را برای دیگر چاه‌های این میدان نیز می‌تواند داشته باشد همچنین محققان می‌توانند از این الگوریتم به منظور پیش‌بینی فشار منفذی برای سایر میدان‌ها و همچنین به دست آوردن پارامترهای دیگر استفاده نمایند.

جدول ۹: تعمیم پیش‌بینی فشار منفذی با استفاده از الگوریتم درخت تصمیم (آموزش دیده شده با استفاده از اطلاعات مربوط به دو چاه #A و #B) برای اطلاعات مربوط به چاه #C (۹۳۴ داده ست).

مدل‌ها	ARE	AARE	STD	MSE	RMSE	R <sup>2</sup>
واحد‌ها	(%)	(%)	(Psi)	(Psi)	(Psi)	-
الگوریتم درخت تصمیم	-0.001	0.237	17.157	2.944E+02	17.1573	0.9953



شکل ۸: نمودار متقاطع به منظور تعمیم پیش‌بینی فشار منفذی برای الگوریتم درخت تصمیم (آموزش دیده شده با استفاده از داده‌های مربوط به دو چاه #A و #B) برای اطلاعات مربوط به چاه #C.

## ۶- نتیجه‌گیری

این مطالعه به منظور پیش‌بینی فشار منفذی، از چهار الگوریتم جدید هوش مصنوعی از جمله الگوریتم جنگل تصادفی، الگوریتم رگرسیون بردار پشتیبان، الگوریتم شبکه عصبی مصنوعی و الگوریتم درخت تصمیم بهره برده است که داده‌های مورد استفاده، ۲۸۲۷ مجموعه داده‌ای مربوط به سه چاه #A، #B و #C واقع در یکی از میدان‌ها نفتی جنوب غرب ایران می‌باشد. پس از بررسی انتخاب ویژگی، بهترین ترکیب اطلاعات استفاده شده شامل ۹ متغیر LLS، CGR، Vp، Vs، SGR، NPHI، PEF، ILD و RHOB می‌باشد. پس از بررسی نتایج مشخص گردید الگوریتم درخت تصمیم دارای دقت عملکرد بسیار بالاتری از سه الگوریتم پیشنهادی دیگر دارد (دقت عملکرد برای کل مجموعه داده‌ها شامل  $R^2=0.9985$  و  $RMSE=14.460\text{psi}$ ). از جمله مزیت‌های این الگوریتم، ارائه بهترین نتیجه بدون نیاز به دانش آماری، جدا کردن داده‌های غیرضروری، آماده‌سازی داده‌ها در زمان کوتاه و کاهش خطای نسبی با یافتن گره اصلی تصمیم گیر است. این قابلیت‌ها سبب شده است که این الگوریتم جایگاهی خاص در بین الگوریتم‌های سری رگرسیون داشته باشد. در نهایت با تعمیم این الگوریتم برای چاه #C، مشخص گردید که این الگوریتم قابلیت استفاده برای چاه‌های دیگر این میدان را نیز دارد.

## در دسترس بودن داده‌ها

از آنجایی که داده‌های مورد استفاده در مطالعه محرمانه هستند، نویسندگان نمی‌توانند آن‌ها را به صورت عمومی و یا به مجله ارسال یا به اشتراک بگذارند.

## مراجع

- [1] Hu L, Deng J, Zhu H, Lin H, Chen Z, Deng F, et al. A new pore pressure prediction method-back propagation artificial neural network. *Electron J Geotech Eng* 2013;18:4093-107. doi: <http://www.ejge.com/2013/Ppr2013.371mlr.pdf>.
- [2] Yu H, Chen G, Gu H. A machine learning methodology for multivariate pore-pressure prediction. *Computers & Geosciences* 2020;143:104548. doi: <https://doi.org/10.1016/j.cageo.2020.104548>.
- [3] Bahmaei Z, Hosseini E. Pore pressure prediction using seismic velocity modeling: case study, Sefid-Zakhor gas field in Southern Iran. *Journal of Petroleum Exploration and Production Technology* 2020;10(3):1051-62. doi: <https://link.springer.com/article/10.1007/s13202-019-00818-y>.
- [4] Zhang Y, Lv D, Wang Y, Liu H, Song G, Gao J. Geological characteristics and abnormal pore pressure prediction in shale oil formations of the Dongying depression, China. *Energy Science & Engineering* 2020;8(6):1962-79. doi: <https://doi.org/10.1002/ese3.641>.
- [5] Mousavipour F, Riahi MA, Moghanloo HGJJoPE, Technology P. Prediction of in situ stresses, mud window and overpressure zone using well logs in South Pars field. 2020;10(5):1869-79. doi: <https://doi.org/10.2118/189665-PA>.
- [6] Ahedor MK-N, Anumah P, Sarkodie-Kyeremeh J. Post-Drill Pore Pressure and Fracture Gradient Analyses of Y-Field in the Offshore Tano Basin of Ghana. *OnePetro*; 2020. doi: <https://doi.org/10.2118/203659-MS>.
- [7] Mahetaji M, Brahma J, Sircar A. Pre-drill pore pressure prediction and safe well design on the top of Tulumura anticline, Tripura, India: a comparative study. *Journal of Petroleum Exploration and Production Technology* 2020;10(3):1021-49. doi: <https://link.springer.com/article/10.1007/s13202-019-00816-0>.
- [8] Oloruntobi O, Falugba O, Ekanem-Attah O, Awa C, Butt S. The Niger Delta basin fracture pressure prediction. *Environmental Earth Sciences* 2020;79(13):1-11. doi: <https://link.springer.com/article/10.1007/s12665-020-09081-5>.
- [9] Darvishpour A, Seifabad MC, Wood DA, Ghorbani H. Wellbore stability analysis to determine the safe mud weight window for sandstone layers. *Petroleum Exploration and Development* 2019;46(5):1031-8. doi: [https://doi.org/10.1016/S1876-3804\(19\)60260-0](https://doi.org/10.1016/S1876-3804(19)60260-0).
- [10] Li S, George J, Purdy C. Pore-pressure and wellbore-stability prediction to increase drilling efficiency. *Journal of Petroleum Technology* 2012;64(02):98-101. doi: <https://doi.org/10.2118/144717-JPT>.
- [11] Benemaran RS, Esmacili-Falak M. Optimization of cost and mechanical properties of concrete with admixtures using MARS and PSO. *Computers and Concrete, An International Journal* 2020;26(4):309-16. doi: [https://doi.org/10.1061/\(ASCE\)CR.1943-5495.0000188](https://doi.org/10.1061/(ASCE)CR.1943-5495.0000188).
- [12] Zhu W, Huang L, Mao L, Esmacili-Falak M. Predicting the uniaxial compressive strength of oil palm shell lightweight aggregate concrete using artificial intelligence-based algorithms. *Structural Concrete* 2021. doi: <https://doi.org/10.1002/suco.202100656>.
- [13] Hottmann CE, Johnson RK. Estimation of formation pressures from log-derived shale properties. *Journal of Petroleum Technology* 1965;17(06):717-22. doi: <https://doi.org/10.2118/1110-PA>.
- [14] Bingham G. A new approach to interpreting rock drillability. *TECHNICAL MANUAL REPRINT, OIL AND GAS JOURNAL*, 1965 93 P 1965. doi:
- [15] Jordan JR, Shirley OJ. Application of drilling performance data to overpressure detection. *Journal of Petroleum Technology* 1966;18(11):1387-94. doi: <https://doi.org/10.2118/1407-PA>.
- [16] Terzaghi K, Peck RB, Mesri G. *Soil mechanics in engineering practice*. John Wiley & Sons; 1996.
- [17] Biot MA, Willis DG. The elastic coefficients of the theory of consolidation. 1957. doi: <https://doi.org/10.1115/1.4011606>.
- [18] Eaton BA. The equation for geopressure prediction from well logs. *OnePetro*; 1975. doi: <https://doi.org/10.2118/5544-MS>.
- [19] Bowers GL. Pore pressure estimation from velocity data: Accounting for overpressure mechanisms besides undercompaction. *SPE Drilling & Completion* 1995;10(02):89-95. doi: <https://doi.org/10.2118/27488-PA>.

- [20] Yoshida C, Ikeda S, Eaton BA. An investigative study of recent technologies used for prediction, detection, and evaluation of abnormal formation pressure and fracture pressure in North and South America. *OnePetro*; 1996. doi: <https://doi.org/10.2118/36381-MS>.
- [21] Shen Y, Luan G, Zhang H, Liu Q, Zhang J, Ge H. Novel method for calculating the effective stress coefficient in a tight sandstone reservoir. *KSCE Journal of Civil Engineering* 2017;21(6):2467. doi: 10.1007/s12205-016-0514-5.
- [22] Farsi M, Mohamadian N, Ghorbani H, Wood DA, Davoodi S, Moghadasi J, et al. Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms. *Natural Resources Research* 2021;1-27. doi: <https://link.springer.com/article/10.1007/s11053-021-09852-2>.
- [23] Abad ARB, Mousavi S, Mohamadian N, Wood DA, Ghorbani H, Davoodi S, et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs. *Journal of Natural Gas Science and Engineering* 2021;95:104210. doi: <https://doi.org/10.1016/j.jngse.2021.104210>.
- [24] Naveshki M, Naghiei A, Soltani Tehrani P, Ahmadi Alvar M, Ghorbani H, Mohamadian N, et al. Prediction of bubble point pressure using new hybrid computational intelligence models. *Journal of Chemical and Petroleum Engineering* 2021. doi: 10.22059/JCHPE.2021.314719.1341.
- [25] Hazbeh O, Ahmadi Alvar M, Aghdam K-y, Ghorbani H, Mohamadian N, Moghadasi J. Hybrid computing models to predict oil formation volume factor using multilayer perceptron algorithm. *Journal of Petroleum and Mining Engineering* 2021;14-27. doi: 10.21608/JPME.2021.52149.1062.
- [26] Wang Y, Ma H, Fu W. Formation pressure prediction based on hybrid genetic algorithm. *IEEE*; 2010:2535-8. doi: 10.1109/ICOSP.2010.5656925.
- [27] Abidin MH. Pore Pressure Estimation Using Artificial Neural Network. 2014. doi: <http://utpedia.utp.edu.my/id/eprint/14317>.
- [28] Haris A, Sitorus RJ, Riyanto A. Pore pressure prediction using probabilistic neural network: case study of South Sumatra Basin. *IOP Conference Series: Earth and Environmental Science* 2017;62:012021. doi: 10.1088/1755-1315/62/1/012021.
- [29] Rashidi M, Asadi A. An Artificial Intelligence Approach in Estimation of Formation Pore Pressure by Critical Drilling Data. *OnePetro*; 2018. doi:
- [30] Karmakar M, Maiti S. Short term memory efficient pore pressure prediction via Bayesian neural networks at Bering Sea slope of IODP expedition 323. *Measurement* 2019;135:852-68. doi: <https://onepetro.org/ARMAUSRMS/proceedings-abstract/ARMA18/All-ARMA18/ARMA-2018-1098/124075>.
- [31] Abdelaal A, Elkatatny S, Abdulraheem A. Data-Driven Modeling Approach for Pore Pressure Gradient Prediction while Drilling from Drilling Parameters. *ACS omega* 2021. doi: <https://doi.org/10.1021/acsomega.1c01340>.
- [32] Legg S, Hutter M. A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and applications* 2007;157:17. doi:
- [33] Russell S, Norvig P. Artificial intelligence: a modern approach. 2002. doi: <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/27702.pdf>.
- [34] Poole D, Mackworth A, Goebel R. *Computational Intelligence*. 1998. doi:
- [35] Nie G, Rowe W, Zhang L, Tian Y, Shi Y. Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications* 2011;38(12):15273-85. doi: <https://doi.org/10.1016/j.eswa.2011.06.028>.
- [36] Tsai C-F, Chiou Y-J. Earnings management prediction: A pilot study of combining neural networks and decision trees. *Expert systems with applications* 2009;36(3):7183-91. doi: <https://doi.org/10.1016/j.eswa.2008.09.025>.
- [37] Lorena AC, de Carvalho AC. Protein cellular localization prediction with support vector machines and decision trees. *Computers in Biology and Medicine* 2007;37(2):115-25. doi: <https://doi.org/10.1016/j.compbiomed.2006.01.003>.
- [38] Osei-Bryson K-M. Evaluation of decision trees: a multi-criteria approach. *Computers & Operations Research* 2004;31(11):1933-45. doi: [https://doi.org/10.1016/S0305-0548\(03\)00156-4](https://doi.org/10.1016/S0305-0548(03)00156-4).
- [39] Fakhari A, Moghadam AME. Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval. *Applied Soft Computing* 2013;13(2):1292-302. doi: <https://doi.org/10.1016/j.asoc.2012.10.019>.
- [40] Liu J, Sui C, Deng D, Wang J, Feng B, Liu W, et al. Representing conditional preference by boosted regression trees for recommendation. *Information Sciences* 2016;327:1-20. doi: <https://doi.org/10.1016/j.ins.2015.08.001>.
- [41] Ahmad MS, Adnan SM, Zaidi S, Bhargava P. A novel support vector regression (SVR) model for the prediction of splice strength of the unconfined beam specimens. *Construction and Building Materials* 2020;248:118475. doi: <https://doi.org/10.1016/j.conbuildmat.2020.118475>.
- [42] Rui J, Zhang H, Zhang D, Han F, Guo Q. Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization. *Journal of Petroleum Science and Engineering* 2019;180:699-706. doi: <https://doi.org/10.1016/j.petrol.2019.06.014>.
- [43] Shao M, Wang X, Bu Z, Chen X, Wang Y. Prediction of energy consumption in hotel buildings via support vector machines. *Sustainable Cities and Society* 2020;57:102128. doi: <https://doi.org/10.1016/j.scs.2020.102128>.

- [44] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and computing* 2004;14(3):199-222. doi: <https://link.springer.com/article/10.1023/B:STCO.0000035301.49549.88>.
- [45] Vapnik V. *The nature of statistical learning theory*. Springer science & business media; 2013.
- [46] Hashemitaheri M, Mekarthy SMR, Cherukuri H. Prediction of specific cutting forces and maximum tool temperatures in orthogonal machining by support vector and Gaussian process regression methods. *Procedia Manufacturing* 2020;48:1000-8. doi: <https://doi.org/10.1016/j.promfg.2020.05.139>.
- [47] Zhou X, Lu P, Zheng Z, Tolliver D, Keramati A. Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety* 2020;200:106931. doi: <https://doi.org/10.1016/j.ress.2020.106931>.
- [48] Grape S, Branger E, Elter Z, Balkesthahl LP. Determination of spent nuclear fuel parameters using modelled signatures from non-destructive assay and Random Forest regression. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 2020;969:163979. doi: <https://doi.org/10.1016/j.nima.2020.163979>.
- [49] Ahmad MW, Reynolds J, Rezguy Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of cleaner production* 2018;203:810-21. doi: <https://doi.org/10.1016/j.jclepro.2018.08.207>.
- [50] Ahmad MW, Mourshed M, Rezguy Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings* 2017;147:77-89. doi: <https://doi.org/10.1016/j.enbuild.2017.04.038>.
- [51] Shahbaz M, Taqvi SA, Loy ACM, Inayat A, Uddin F, Bokhari A, et al. Artificial neural network approach for the steam gasification of palm oil waste using bottom ash and CaO. *Renewable Energy* 2019;132:243-54. doi: <https://doi.org/10.1016/j.renene.2018.07.142>.
- [52] Kartal F, Özveren U. A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. *Energy* 2020;209:118457. doi: <https://doi.org/10.1016/j.energy.2020.118457>.
- [53] Abad ARB, Ghorbani H, Mohamadian N, Davoodi S, Mehrad M, Aghdam SK-y, et al. Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields. *Fuel* 2022;308:121872. doi: <https://doi.org/10.1016/j.fuel.2021.121872>.
- [54] Wahab MNA, Nefti-Meziani S, Atyabi A. A comprehensive review of swarm optimization algorithms. *PloS one* 2015;10(5):1-36. doi: <https://doi.org/10.1371/journal.pone.0122827>.
- [55] Chandrashekar G, Sahin F. A survey on feature selection methods. *Computers and Electrical Engineering* 2014;40(1):16-28. doi: <https://doi.org/10.1016/j.compeleceng.2013.11.024>.
- [56] Al-Mudhafar WJ. Bayesian and LASSO regressions for comparative permeability modeling of sandstone reservoirs. *Natural Resources Research* 2019;28(1):47-62. doi: <https://link.springer.com/article/10.1007/s11053-018-9370-y>.
- [57] Jain A, Zongker D. Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence* 1997;19(2):153-8. doi: <https://doi.org/10.1109/34.574797>.
- [58] Fu G-H, Wu Y-J, Zong M-J, Pan J. Hellinger distance-based stable sparse feature selection for high-dimensional class-imbalanced data. *BMC bioinformatics* 2020;21(1):1-14. doi: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3411-3>.
- [59] Salehi M, Farhadi S, Moieni A, Safaie N, Ahmadi HJFips. Mathematical modeling of growth and paclitaxel biosynthesis in *Corylus avellana* cell culture responding to fungal elicitors using multilayer perceptron-genetic algorithm. 2020;11. doi: 10.3389/fpls.2020.01148.
- [60] Jotheeswaran J, Koteeswaran S. Sentiment Polarity Classification Using Conjure of Genetic Algorithm and Differential Evolution Methods for Optimized Feature Selection. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* 2020;13(6):1284-91. doi: <https://doi.org/10.2174/2213275911666180904110105>.