انتخاب مته حفاری بهینه با استفاده از الگوریتم‌های داده‌کاوی-مطالعهٔ موردی

هادی فتاحی*، یونس افشاری؛ دانشگاه صنعتی اراک، دانشکدهٔ مهندسی علوم زمین، گروه مهندسی ژئومکانیک

تاریخ: دریافت ۹۸/۰۳/۱۶ پذیرش ۹۸/۰۶/۲۵

چکیده

انتخاب بهترین مته در شرایط پیچیده حفاری متناظر بـا آن، یکی از مهم‌ترین موضوعاتی است که در حوزهٔ حفاری وجود دارد. زیرا با وجود این‌که قیمت مته ۲ تا ۳ درصد هزینه‌های تکمیل یک چاه را در بر می‌گیرد، اما بر ۷۵ درصد هزینه‌های کلـی حفـاری به‌طور غیرمستقیم تأثیرگذار است. در این تحقیق به مدل‌سازی انتخاب مته حفاری بهینه با استفاده از چاه نمودارهای (لاگ) مختلف ۷ چاه نفتی موجود در منطقه‌ای در ترکیه پرداخته شد. برای مدل‌سازی از روش‌های داده‌کاوی شامل درخت تصمیم، قوانین انجمنی، احتمال بیز، مبتنی بر تشابه و سیستم استنتاجی نروفازی تطبیقی استفاده شد. بدین‌ترتیب که از داده‌های شش چاه به‌عنوان آموزش مدل‌ها و داده‌های یک چاه دیگر به‌عنوان داده‌های آزمون برای ارزیابی صحت و دقت مدل‌ها استفاده شد. در نهایت نتایج مدل‌های مختلف در کنار یک‌دیگر مقایسه و تحلیل شد. نتایج نشان داد مدل ایجاد شده به‌وسیلهٔ سیستم استنتاجی نروفازی تطبیقی با اختلاف معناداری از مدل‌های ایجاد شده به‌وسیلهٔ سایر روش‌ها کاراتر و دقیق‌تر است. اما بدین معنی نیست که سایر روش‌ها کارا نیستند بلکه تحلیل نتایج نشان می‌دهد دیگر روش‌ها نیز می‌توانند مدلی هرچند در کیفیتی پایین‌تر از مدل سیستم استنتاجی نروفازی تطبیقی اما سودمند و قابل اعتماد ایجاد کنند.

واژه‌های کلیدی: انتخاب مته حفاری، چاه‌نمودارهای (لاگ) مختلف، سیستم استنتاجی نروفازی تطبیقی، درخت تصمیم، قوانین انجمنی، احتمال بیز، مبتنی بر تشابه

مقدمه

مته حفاری اصلی‌ترین ابزار اسـتفاده شدهٔ مهنـدس حفـار است و انتخاب بهترین مته در شرایط حفاری متناظر بـا آن، یکی از اساسی‌ترین مشکلاتی است که حفاران با آن مواجه

*نویسنده مسئول h.fattahi@arakut.ac.ir

هستند [۱]. با وجود این‌که قیمت مته ۲ تا ۳ درصد هزینه‌های تکمیل یک چاه را در بر می‌گیرد، اما بر ۷۵ درصد هزینه‌های کلی حفاری به‌طور غیرمستقیم، که شامل ۴۵ درصد هزینه تکمیل یک چاه است، تأثیرگذار است [۲]. مته‌ها بر اساس جنس، نوع برش دندانه‌ها، محل قرارگیری نازل‌های سیال حفاری و دیگر پارامترها بسیار با هم متفاوت هستند. از این‌رو، انتخاب متهٔ مناسب برای شرایط حفاری خاص، نیازمند ارزیابی عوامل متعددی است. البته اگر تمامی‌پارامترهایی که در انتخاب مته تأثیرگذارند، بررسی شوند، انتخاب مته بسیار پیچیده می‌شود. بنابراین در هر یک از روش‌های انتخاب مته، تنها پارامترهای محدودی را ارزیابی می‌کنند. تا همین اواخر، انتخاب مناسب‌ترین مته از طریق آزمون و خطا صورت می‌گرفت. باید توجه داشت که انتخاب مته نامناسب گاه باعث صرف هزینه‌های بسیار در عملیات حفاری می‌شود. در زمینهٔ انتخاب بهینه مته حفاری، روش‌هایی وجود دارد که در بخش بعدی مرور می‌شود.

روش‌های قدیمی‌شامل هزینهٔ حفاری واحد طول، انرژی ویژه و قابلیت حفاری سازند هستند که از جمله نقاط ضعف این روش‌ها محدود بودن آن‌ها به استفاده از داده‌های خاصی مانند زمان حفاری، مدت زمان زیاد و کم کردن متعلقات حفاری در چاه، قیمت مته حفاری و کار نیروی دورانی لوله حفاری است. از این‌رو، در اکثر اوقات این اطلاعات در دسترس نیست و یا جمع‌آوری این نوع داده‌ها با هزینه زیاد و خطا همراه است. به‌علاوه این‌که استفاده از روش‌های مذکور در صورت وجود یک پایگاه حجیم داده بسیار زمان‌بر و طاقت‌فرسا است. وابسته بودن به عوامل غیرمرتبط در کارایی مته، عدم توانایی در ارزیابی مته، در نظر نگرفتن پارامترهای مربوط به سازند و هم‌چنین نیاز به اندازه‌گیری گشتاور در تمام طول چاه از دیگر ضعف‌های این روش‌ها است. در تحقیقات جدیدتر نیز از بین روش‌های نوین تنها از روش شبکه عصبی مصنوعی استفاده شده است که دارای ضعف‌هایی است از جمله آنها این است که این روش به‌عنوان یک جعبهٔ سیاه عمل می‌کند و اطلاعاتی در رابطه با دانش موجود بین ورودی‌ها و خروجی‌ها نمی‌دهد. امروزه تکنیک‌های داده‌کاوی به‌عنوان زیرمجموعه‌ای از روش‌های هوش محاسباتی، پایگاه‌ها و مجموعه‌های حجیم داده را برای کشف و استخراج دانش تحلیل می‌کنند. در سال‌های اخیر داده کاوی، با توجه به دسترسی گسترده به مقادیر بسیار زیاد داده و نیاز به تبدیل چنین داده‌هایی به اطلاعات و دانش مفید توجه زیادی را

به‌خود جلب کرده است [۳]. الگوریتم‌های بسیار زیادی برای پیدا کردن الگوهای تکراری وجود دارند که در این مقاله مهم‌ترین آن‌ها برای انتخاب بهینه مته حفاری به‌کار گرفته شده است.

مروری بر کارهای گذشته

از گذشته محققان از روش‌های مختلفی برای انتخاب مته استفاده کرده‌اند. تلاش آن‌ها سبب ارائه روش‌هایی مانند هزینهٔ حفاری سازند، انرژی ویژه، روش قابلیت حفاری سازند و روش‌های هوش محاسباتی شد که در ادامه به هریک از آن‌ها به اختصار اشاره می‌شود [٤]، [٥].

۱.روش هزینه حفاری

انتخاب مته بر مبنای هزینهٔ حفاری، بر اساس فاکتورهایی مانند هزینه دکل، قیمت مته و نرخ نفوذ برای واحد طول انجام می‌شود [٦]. بیش‌ترین و معمول‌ترین کاربرد روش هزینه حفاری، ارزیابی کارایی مته رانده شده است. هزینهٔ حفاری واحد طول از یک چاه برابر است با مجموع قیمت مته، هزینه تعویض مته و هزینه عملیاتی دکل حفاری در زمان و بازهٔ حفاری موردنظر. اگر هزینه راندن مته بر طول حفاری شده تقسیم شود، هزینهٔ حفاری واحد طول بازه حفاری شده به‌دست می‌آید. هزینهٔ حفاری واحد طول یک مته، در صورت انتخاب صحیح وزن روی مته و سرعت دوران، حداقل می‌شود. وزن روی مته و سرعت دوران فقط بر دو پارامتر تأثیر می‌گذارند که عبارت است از: هزینهٔ حفاری در حین دوران مته ($Cr.tb$) و طول حفاری شده (ΔD). از طرفی قیمت مته و هزینه زیاد و کم کردن رشته حفاری و تعویض مته ثابت است [۷]. کل زمانی که برای حفاری بازهٔ مورد نظر (ΔD) لازم است برابر مجموع زمان حفاری (tb)، زمان توقف مته (tc)، زمان زیاد و کم کردن رشته حفاری و زمان تعویض مته (tt) است. در نهایت رابطهٔ هزینه حفاری به‌صورت (۱) ارائه می‌شود [۸]:

$$C_f = \frac{C_b + C_r\left(t_b + t_c + t_t\right)}{\Delta D} \qquad (۱)$$

که در آن Cf هزینه حفاری واحد طول، Cb قیمت مته و Cr هزینه‌های ثابت عملیاتی دکل حفاری در واحد زمان و مستقل از دیگر عوامل ارزیابی شده است [۹].

میسون[1] در پژوهشی با استفاده از اطلاعات حفاری‌های انجام شده در سازندهای ماسه

1. Mason

سنگی نفتی در شرق تگزاس با استفاده از روش هزینه حفاری واحد طول، مته‌های مناسب برای اعماق مختلف را مشخص کرد [۱۰]. در همین راستا، فرناندو[1] و همکاران از اطلاعات حفاری‌های انجام شده در برزیل برای انتخاب مته بهینه استفاده کردند. آنها با استفاده از معادله روش هزینه حفاری، هزینه حفاری برای هر فوت از سازند را محاسبه کردند. آنها توانستند مته‌های بهینه را برای حفاری در اعماق مختلف چاه‌ها مشخص کنند [۱۱]. از نتایج این تحقیق این نکته حاصل شد که استفاده از تنها یک مته برای حفاری یک سازند مناسب نیست و برای حفاری در یک سازند، بر اساس عمق و جنس لایه باید از مته‌های مختلف استفاده کرد. فالکائو[2] و همکاران از اطلاعات ۹۷ حلقه چاه نفتی در برزیل برای یافتن مته بهینه حفاری از روش هزینه حفاری واحد طول استفاده کردند. آنها نشان دادند مته نوع p با کم‌ترین هزینه حفاری بهترین مته حفاری برای عمق ۲۰۰ فوتی است [۱۲]. برتری این پژوهش نسبت به پژوهش‌های قبلی در این بود که محققان توانسته بودند برای هر عمق از حفاری مته‌ای مناسب را انتخاب کنند. این دست آورد سبب شد که برای حفاری یک طول مشخص از یک چاه، مجموعه مته‌های بهینه در آن طول بررسی شود. با در نظر گرفتن هزینه تعویض و بالا و پایین بردن مته به‌جای استفاده از چند مته برای حفاری از یک متهٔ مشخص بهینه که سبب کاهش هزینه حفاری می‌شود، می‌توان استفاده کرد. مک‌آدامز[3] و همکاران که روی حفاری‌های انجام شده در سازندهای ماسه سنگی نفتی غرب تگزاس انجام شد، دریافتند که پارامترهای مؤثر بر حفاری (وزن روی مته، سرعت دوران مته، هیدرولیک مته وغیره) روی هزینهٔ حفاری بسیار تأثیرگذارند [۱۳]. آنها نشان دادند که طراحی مته‌ها بر اساس بهترین عملکرد سبب می‌شود هزینه حفاری به شدت کاهش یافته و راندمان حفاری افزایش یابد. به‌طور مثال آنها نشان دادند جایگزینی مته‌ها با سرعت دوران بسیار زیاد به‌جای مته‌های با سرعت دوران کم سبب افزایش نرخ نفوذ شده که نهایتاً هزینه حفاری را کاهش می‌دهد. برای رسیدن به نرخ نفوذ مطلوب باید مته متناسب با افزایش سرعت دوران بازطراحی شود. در پایان آنها نتیجه گرفتند مته‌ای که هزینهٔ حفاری واحد طول آن کم باشد،

1. Fernando
3. Falcao
4. *McAdams*

پارامترهای آن (وزن روی مته، سرعت دوران مته، هیدرولیک مته و غیره) به‌نحو مطلوبی در کنار یک‌دیگر کار می‌کنند. پس می‌توان برای انتخاب مته بهینه تنها هزینهٔ حفاری مته را مد نظر قرار داد. اطلاعات به‌دست آمده از عملیات حفاری نشان داد مته با کم‌ترین هزینه حفاری توانسته بهترین راندمان حفاری را در منطقهٔ بررسی شده ایجاد کند. ایکسیو[1] و همکاران اقدام به انتخاب مته حفاری بهینه با استفاده از روش هزینه حفاری واحد طول کردند [۱۴]. آنها نشان دادند بر خلاف پژوهش‌های گذشته تنها بررسی مته‌های استفاده شده در حفاری برای انتخاب مته بهینه نمی‌تواند صحیح باشد زیرا این مقایسه نیازمند گرفتن اطلاعات از عملیات حفاری است. علاوه بر این شاید بهترین مته اصلاً در حفاری منطقه استفاده نشده باشد. به‌همین جهت در این تحقیق سعی شد ارتباط هزینه حفاری با متغیرهای قابل اندازه‌گیری (بدون درنظرگرفتن اطلاعات حاصل از عملیات حفاری) بررسی شود. بر همین اساس روش هزینه حفاری و پارامترهای مؤثر بر آن بازبینی شد و در نهایت رابطهٔ (۲) به‌دست آمد:.

$$CPUD_m = \frac{C_r}{ROP_m} + \frac{C_r}{ROP_m \times t_{bm}} + \frac{C_r \times t_t}{ROP_m \times t_{bm}} \qquad (۲)$$

با توجه به رابطهٔ (۲)، پارامترهای نرخ نفوذ و مدت زمان حفاری مته (tb) بیش‌ترین تأثیر را روی هزینه حفاری دارند. برای مدل‌سازی نرخ نفوذ و مدت زمان حفاری مته دو مدل به‌وسیلهٔ محققان پیشنهاد شد. برای بررسی صحت مدل پیشنهادی اول (مدل نرخ نفوذ) هم‌بستگی بین نرخ نفوذ واقعی و نرخ نفوذ پیشنهادی مدل بررسی شد. هم‌بستگی مناسب ۰/۷۰ نشان از توانایی زیاد مدل در تخمین نرخ نفوذ در منطقهٔ بررسی شده دارد. در این مدل‌سازی نشان داده شد که مدت زمان حفاری مته و τb ارتباط مستقیمی با هم دارند. در نهایت آنها با استفاده از مدل نرخ نفوذ و مدت زمان حفاری مته، هزینهٔ حفاری واحد طول را برای مته‌های مختلف محاسبه کرده و متهٔ بهینه را بر اساس کم‌ترین هزینه حفاری انتخاب کردند.

۲. روش انرژی ویژه

مفهوم انرژی ویژه در حفاری سنگ اولین بار به‌وسیلهٔ تیله[2] [۹] به‌عنوان شاخصی برای اندازه‌گیری کارآیی مکانیکی کارهای انجام شده روی سنگ، پیشنهاد شد. انرژی ویژه به‌عنوان

1. Xu
2. Teale

انرژی مورد نیاز برای حفر واحد حجم سنگ معرفی می‌شود. این مفهوم تاکنون به‌طور بسیار گسترده در پژوهش‌های انجام شده روی سنگ، به‌عنوان شاخص کارآیی و هم به‌عنوان مقیاس قابلیت حفاری استفاده شده است. در حفاری دورانی کار انجام شده را می‌توان به دو بخش کار انجام شده به‌وسیلهٔ نیروی محوری (وزن روی مته W) و کار انجام شده به‌وسیلهٔ مؤلفهٔ دورانی (گشتاور T) تقسیم‌بندی کرد. اگر سرعت دوران N، سطح درگیر مته A، نرخ نفوذ R و مته به اندازه Y پیش‌روی کند، کار انجام شده به‌وسیلهٔ نیروی محوری WY خواهد بود. کار انجام شده به‌وسیلهٔ نیروی مماسی در هر دوران کامل و با پیش‌روی مته به اندازه Y، برابر می‌شود با [۹]:

$$W_1 = 2\pi T\left(\frac{T}{R}\right)Y \qquad (۲)$$

بنابراین کل کار انجام شده با مته $(WTottal)$ با پیش‌روی به اندازهٔ Y و با صرف نظر کردن از انرژی از دست‌رفته یا صرف نظر از کارهایی غیر از حفاری، به‌صورت (۳) نوشته می‌شود:

$$W_{total} = WY + 2\pi N\left(\frac{T}{R}\right)Y \qquad (۳)$$

با توجه به این‌که حجم سنگ جدا شده برابر AY است، انرژی ویژه (SE) برابر می‌شود با:

$$SE = \frac{W}{A} + \frac{2\pi NT}{AR} \qquad (٤)$$

$$SE = \frac{4WR + 8\pi NT}{\pi D^2 R} \qquad (۵)$$

لازم به ذکر است که در معادلات مذکور D قطر مته است [۱۵].

در این زمینه تیله طی پژوهشی، انرژی لازم برای استخراج حجم واحد سنگ (انرژی ویژه) و میزان نیروی لازم برای خرد کردن سنگ‌ها را با یک‌دیگر مقایسه کرد و در نهایت به این نتیجه رسید که پارامترهای روش انرژی ویژه هم‌بستگی زیادی با قدرت خرد کردن و نفوذ در سنگ دارند. وی متوجه شد که قدرت خرد کردن سنگ همیشه مطلق نیست و بسیار وابسته به پارامترهای ژئومکانیکی سازند است. او بر اساس یافته‌های مذکور نشان داد که انرژی ویژه لازم برای خرد کردن سنگ با پارامترهای نظیر سرعت دوران مته، سرعت نفوذ و هندسه مته رابطهٔ تنگاتنگی دارد. پس می‌توان از روش انرژی ویژه برای پیشنهاد سرمته‌های حفاری مناسب استفاده کرد [۹].

رابیا[1] توانست انرژی ویژه حفاری را برای چند کلاس از مته‌های استفاده شده در منطقهٔ نفتی عربستان بر اساس عمق محاسبه کند. به‌طور نمونه وی انرژی ویژه لازم برای حفاری با مته‌های *J22*، *F2* و *J3* از عمق۲۰۰۰ تا ۹۰۰۰ فوتی را محاسبه کرد. وی با توجه به این‌که مته بهینه باید کم‌ترین انرژی را برای برش سنگ مصرف کند، مته *J22* را برای حفاری در بازهٔ ۲۰۰۰ تا ۹۰۰۰ فوتی انتخاب کرد. او با مشخص کردن جنس سازندهای موجود در محل حفاری، اعماق قرارگیری هر سازند در چاه و مشخص کردن مته‌های بهینه برای هر حفاری در طول چاه توانست مته‌های مناسب برای حفاری در هر سازند را مشخص کند [۱۶]. خادم عباس[2] برای انتخاب مته حفاری در دو منطقهٔ کرکوک و بای حسن واقع در شمال عراق از روش انرژی ویژه حفاری برای انتخاب مته استفاده کرد. وی تأثیر انرژی ویژه نسبت به عمق را برای تمام مته‌های استفاده شده در دو منطقه مذکور مشخص کرد. وی با دانستن این نکته که مته بهینه کم‌ترین انرژی ویژه را برای برش سنگ مصرف می‌کند در نهایت بهترین مته برای حفاری در اعماق مختلف را بر اساس کم‌ترین انرژی ویژه برای دو منطقه مذکور مشخص کرد [۱۷]. بوره[3] و همکاران (۲۰۱۵)، در حوضه‌های نفتی هند، نشان دادند که تنها بررسی پارامترهای مؤثر در محاسبهٔ انرژی ویژه برای هر مته نمی‌تواند شاخص مناسبی برای انتخاب مته بهینه باشد و برای انتخاب مته بهینه باید در کنار روش انرژی ویژه عوامل دیگری مانند جنس سازند و پارامترهای اقتصادی نیز مورد توجه قرار گیرند. آنها نشان دادند برای حفاری در یک سازند سخت، انرژی بیش‌تری نسبت به حفاری در یک سازند نرم مورد نیاز است. شاید این موضوع بدیهی به‌نظر برسد اما باید دانست گاهاً انتخاب مته حفاری تنها بر اساس رکوردهای خود مته صورت می‌گیرد و به وضعیت سازند توجهی نمی‌شود [۱۸].

۳. روش قابلیت حفاری سازند

یکی از ساده‌ترین روش‌های انتخاب مته، بر اساس کارکرد مته‌های استفاده شده در همان منطقه است. اما رکورد مته‌ها فقط کارکرد مته را در یک بازهٔ حفاری نشان می‌دهد و نشان‌دهندهٔ این‌که مته در چه شرایطی حفاری کرده است و یا اطلاعات سنگ‌شناسی و یا پارامترهای مقاومتی سازند چه بوده است، نیستند [۱۹]. از سوی دیگر معتبرترین روش پذیرفته شده برای مقایسهٔ

1. Rabia
2. Khadm Abbas
3. Borah

کارکرد مته‌های مختلف، هزینهٔ حفاری واحد طول است. اما این روش نیز به‌دلیل معایبی مانند وابستگی به عوامل غیرمرتبط به کارایی مته، عدم توانایی در ارزیابی مته‌های جدید و هم‌چنین در نظر نگرفتن پارامترهای مربوط به سازند، نمی‌تواند پاسخگوی تمامی‌مشکلات موجود در این زمینه باشد. به‌همین دلیل برخی محققان استفاده از انرژی ویژه حفاری سازند را در کنار هزینهٔ حفاری واحد طول مطرح کرده‌اند که این روش نیز به‌دلیل تأثیر عوامل غیرمرتبط با کارایی مته و هم‌چنین نیاز به اندازه‌گیری گشتاور در تمام طول چاه، که اغلب صورت نمی‌گیرد، با کاستی‌هایی همراه است. به‌دلیل مشکلات ذکر شده در روش‌های قبلی به‌منظور بررسی اثر پارامترهای سازند در کارآیی مته، محققان روش‌هایی را برای کمی‌کردن اثر پارامترهای ژئومکانیکی در انتخاب و کارکرد مته ارائه داده‌اند. این روش‌ها می‌توانند به‌عنوان مکمل روش‌های دیگر انتخاب مته به‌کار گرفته شوند. در مقیاس بزرگ، قابلیت حفاری سازند واکنش یک سازند مشخص را نسبت به حفاری به‌وسیلهٔ انواع مختلف مته، پیش‌بینی می‌کند [۲]. میسون طی پژوهشی نشان داد مدت زمان حفاری با سختی و مقاومت فشاری سازند هم‌بستگی زیادی دارد. وی متوجه شد اگر با توجه به مقدار سختی و مقاومت فشاری سازند متهٔ مناسب را انتخاب کند می‌تواند مدت زمان حفاری را کاهش دهد. وی با استفاده از نگاره‌های صوتی حاصل از سازندهای نفتی شرق تگزاس، سختی لایه‌های موجود را مشخص کرد. در نهایت با استفاده از اطلاعات به‌دست آمده در کنار استفاده از روش هزینه حفاری مته‌های بهینه برای سازندهای موجود در اعماق مختلف مشخص شدند [۱۰]. یوبولدی[1] و همکاران در پژوهشی روی حفاری‌های انجام شده در جنوب ایتالیا بر اساس اندازه‌گیری نیروی لازم برای برش سنگ، یک روش برای انتخاب مته پیشنهاد کردند. آنها نشان دادند می‌توان از نیروی لازم برای برش سنگ به‌عنوان اطلاعاتی برای انتخاب یا طراحی مته بهینه استفاده کرد [٤]. جوانی[2] و همکاران (۲۰۱۵)، تأکید در بررسی و استفاده از پارامترهای زمین‌شناسی و ژئومکانیکی در انتخاب مته مناسب داشتند. آنها بیان داشتند که انتخاب مته حفاری امروزه به دو روش محاسبه انرژی ویژه و یا هزینهٔ حفاری واحد طول انجام می‌گیرد. از جمله کاستی‌های این دو روش، نبود توجه کافی به پارامترهای ژئومکانیکی منطقه است. در این مقاله خروجی دو روش بالا با روش توانایی حفاری سازند، که توجه خاصی به پارامترهای

1. Uboldi
2. Javani

ژئومکانیکی و زمین‌شناسی سازند دارد مقایسه شد. نتایج نشان داد که مقدار انرژی ویژه و هزینهٔ حفاری واحد طول در تشکیلات زمین‌شناسی یک‌سان برای چاه‌های مختلف ممکن است متفاوت باشد که این به‌دلیل تفاوت درخصوصیات زمین‌شناسی و ژئومکانیکی بین چاه-های مختلف یا حتی فواصل مختلف در یک چاه است. آنها نشان دادند با وجود این واقعیت که روش‌های انرژی ویژه و هزینهٔ حفاری واحد طول روش‌های رایج در انتخاب مته حفاری مناسب هستند، عمیقاً نیازمند روش‌های مکمل مانند قابلیت حفاری سازند، به‌منظور استفاده از ویژگی‌های زمین شناسی و ژئومکانیکی در ارزیابی مته حفاری هستند [۲۰].

٤. روش‌های هوش محاسباتی

روش‌های پیشرفته هوش محاسباتی به مجموعه‌ای از شیوه‌های جدید محاسباتی در علوم رایانه، هوش مصنوعی، یادگیری ماشینی و بسیاری از زمینه‌های کاربردی دیگر اطلاق می‌شود. در تمامی‌این زمینه‌ها به بررسی، مدل‌سازی و تجزیهٔ پدیده‌های بسیار پیچیده‌ای نیاز است که شیوه‌های علمی در گذشته به حل آسان، تحلیلی و کامل آنها موفق نبوده‌اند. روش‌های هوش محاسباتی با محور قرار دادن ذهن انسان پیش می‌رود. روش‌های پیشرفته هوش محاسباتی را می‌شود حاصل تلاش‌های جدید علمی‌دانست که مدل‌سازی، تحلیل و در نهایت کنترل سامانه‌های پیچیده را با سهولت و موفقیت زیادتری امکان‌پذیر می‌سازد [۲۱]، [۲۲].

در زمینه استفاده از روش‌های هوش محاسباتی در انتخاب مته حفاری می‌توان به مواردی اشاره کرد از آن جمله: بیلگسو[1] و همکاران در تحقیقی بر اساس اطلاعات حفاری‌های انجام شده در منطقه خاورمیانه و با استفاده از روش شبکه عصبی مصنوعی، مته مناسب برای حفاری را انتخاب کردند. در این روش از پارامترهای سازند صرف نظر شد و تنها پارامترهای حاصل از رکوردهای مته‌های حفاری به‌عنوان داده‌ها ورودی در سه دسته تقسیم‌بندی و استفاده شدند. از مجموعهٔ هر دسته داده ورودی، ۹۰ درصد به‌عنوان داده آموزشی و ۱۰ درصد به‌عنوان داده‌های صحت‌سنجی استفاده شد. بعد از استفاده از شبکهٔ عصبی مصنوعی طراحی شده برای هر دسته از اطلاعات یک مدل برای هر دسته با روش شبکهٔ عصبی مصنوعی پیشنهاد شد. برای بررسی صحت مدل‌ها، سرمته‌های پیشنهادی به‌وسیلهٔ مدل مربوط به هر دسته از اطلاعات در مقابل بهترین مته‌های استفاده شده در محل قرار گرفت و نتایج

1. Bilgesu

نشان داد هم‌بستگی زیاد بین مته‌های پیشنهادی مدل‌ها و بهترین مته‌های استفاده شده در حفاری‌های قبلی وجود داشته است. این تحقیق نشان داد که می‌توان از این مدل‌سازی شبکه عصبی مصنوعی برای حفاری‌های بهینه آینده در محل استفاده کرد [۲۳].

بیلگسو و همکاران (۲۰۰۱)، روش نوینی بر اساس هوش مصنوعی برای انتخاب مته بهینه حفاری پیشنهاد کردند. در این تحقیق از مجموع اطلاعات ثبت شده از حفاری‌های دو منطقه واقع در خاورمیانه به‌عنوان ورودی برای یک شبکه عصبی مصنوعی سه لایه پیش‌رو استفاده شد. آنها در این روش از دو مدل برای پیش‌بینی مته حفاری مناسب استفاده کردند. در مدل اول از پارامترهای اندازه‌گیری شده طی حفاری‌های انجام شده در محل به‌عنوان ورودی برای انتخاب مستقیم مته بهینه استفاده کردند و مقایسهٔ خروجی مدل اول با مته‌های مناسب استفاده شده در محل نشان‌دهندهٔ ضریب هم‌بستگی زیاد (۰/۹۹۷) بین مته‌های پیشنهادی به‌وسیلهٔ مدل و بهترین مته‌های استفاده شده در محل بود. این هم‌بستگی سبب تأیید صحت مدل شد. در مدل دوم از پارامترهای اطلاعات ثبت شده حفاری‌های انجام شده در محل برای پیش بینی مته مناسب بر اساس کم‌ترین هزینه حفاری استفاده شد. بررسی خروجی مدل دوم با مته‌های با کم‌ترین هزینه‌های استفاده شده در محل یک هم‌بستگی بسیار زیاد (۰/۹٦۳) بین این دو را نشان داد که این هم‌بستگی مجدداً مدل پیشنهادی را تأیید کرد. ضریب هم‌بستگی زیاد هر دو مدل نشان داد که می‌توان از این مدل‌ها برای بهینه‌سازی حفاری‌های آینده و انتخاب مته مناسب در منطقه استفاده کرد [۱]. ایلماز[1] و همکاران (۲۰۰۲)، برای انتخاب مته از مدل‌سازی به‌وسیلهٔ شبکهٔ عصبی مصنوعی استفاده کردند. آنها در معماری ساختار شبکه از تلفیق روش‌های شبکه‌های رو به جلو و پس انتشار خطا در جهت انتخاب مته حفاری بهینه در جنوب ترکیه استفاده کردند. هم‌چنین، از اطلاعاتی مانند لاگ صوتی چاه، لاگ گاما، عمق‌های حفاری و دیگر اطلاعات سازند به‌عنوان داده‌های ورودی یک شبکه عصبی مصنوعی پیش‌خور برای پیش‌بینی کد *IADC* مته بهینه استفاده کردند. هم‌بستگی زیاد نتایج مدل با مته‌های استفاده شده در حفاری‌های قبلی، مدل را تأیید کرد و از مدل برای پیش‌بینی بهترین مته‌ها در منطقه استفاده شد [۲٤]. بتایی[2] و همکاران (۲۰۱۰)، ابتدا مدل‌های مختلف

1. Yılmaz
2. Bataee

تجربی پیشنهاد شده به‌وسیلهٔ بورگوین[1]، واران[2] و بینگهام[3] را برای بررسی نرخ نفوذ ارزیابی کردند. نتایج نشان داد که مدل بورگوین کم‌ترین خطا را در پیش‌بینی نرخ نفوذ برای مته‌های مخروطی داشته است. اما در ادامه بتایی و همکاران یک شبکهٔ عصبی مصنوعی سه‌لایه برای پیش‌بینی نرخ نفوذ را برای مقایسه با معادلات تجربی پیشنهادی برای پیش‌بینی نرخ نفوذ پیشنهاد کردند. نتایج این مقایسه نشان داد استفاده از روش شبکه عصبی مصنوعی در صورتی که به‌نحو مناسبی طراحی و تنظیم شود می‌تواند بهتر از معادلات تجربی عمل کند [۲۵]. در تحقیقی دیگر در سال۲۰۱۱ بتایی و همکاران یک شبکه عصبی مصنوعی سه‌لایه با ۸ ورودی (نرخ نفوذ، وزن روی مته، نرخ دوران مته، شدت جریان گل حفاری، فشار گل، عمق حفاری و مقاومت فشاری تک محوری) به‌نحوی تنظیم کردند که خروجی کد مته حفاری باشد. سیستم پیشنهاد شده برای منطقه شادگان ایران استفاده شد و مقایسه نتایج پیشنهاد شده به‌وسیلهٔ سیستم و نتایج حاصل از حفاری در منطقه دقت و صحت مدل را تأیید کرد [۲۶]. عدالت‌خواه[4] و همکاران (۲۰۱۰)، برای انتخاب مته حفاری مناسب از مدل‌سازی با شبکه عصبی مصنوعی استفاده کردند. آنها برای انتخاب بهترین سرمته از دو مدل‌سازی استفاده کردند که مدل‌سازی اول برای دست‌یابی به بهترین مته بر اساس کد *IADC* بود. در مدل دوم از کدهای*IADC* به‌عنوان ورودی به سیستم برای یافتن بهترین نرخ نفوذ (*ROP*) استفاده شد. آنها از مجموعه اطلاعات حاصل شده از گزارش‌های حفاری پارس جنوبی ۶۰ درصد را به‌عنوان داده آموزشی و ۴۰ درصد را به‌عنوان داده‌های دقت‌سنجی استفاده کردند. برای بررسی مدل‌ها، مته‌های پیشنهادی به‌وسیلهٔ مدل‌ها را با بهترین مته‌های استفاده شده در محل حفاری مقایسه کرده‌اند. این بررسی و مقایسه‌ها نشان داد خروجی‌های هر دو مدل با اطلاعات حفاری هم‌بستگی زیادی دارند و می‌توانند از هر دو مدل برای انتخاب کد مته‌های مناسب با نرخ نفوذ مورد نظر در حفاری‌های آیندهٔ منطقه استفاده کرد [۲۷]. مؤمنی[5] و همکاران (۲۰۱۷)، علاوه بر اطلاعات کمی‌حفاری (وزن روی مته، سایز مته و...) از پارامتر جدیدی با نام "ویژگی تصویری مته" برای انتخاب مته استفاده کردند. آنها تصاویر سطح مته‌های

1. Bourgoyne and young
2. Warren
3. Bingham
4. Edalatkhah
5. Momeni

مختلفی را تحت شرایط کنترل شده و یک‌سان تهیه کردند. با استفاده از این تصاویر پارامترهای جدیدی از جمله رابطهٔ بین محور بزرگ و کوچک مته، مساحت سطح مته، زاویهٔ بین محور اصلی مته و سطح تماس و غیره، از مته استخراج شد و این اطلاعات در کنار اطلاعات کمی‌حفاری به‌عنوان اطلاعات ورودی یک شبکهٔ عصبی مصنوعی استفاده شد. آزمایش و تحلیل نتایج مقادیر مناسبی را برای شاخص‌های خطا (میانگین مربعات خطا) نشان داد. کیفیت خوب نتایج نشان داد که مدل امکان استفاده را برای مناطق جدید دارد [۲۸]. مؤمنی و همکاران (۲۰۱۸)، از تلفیق شبکهٔ عصبی مصنوعی و الگوریتم ژنتیک برای انتخاب مته بهینه حفاری در یک منطقه نفتی در جنوب ایران استفاده کردند. آنها معماری ساختار شبکه عصبی مصنوعی را به الگوریتم زنتیک سپردند. در نهایت تلفیق این دو الگوریتم یک شبکهٔ عصبی مصنوعی سه لایه با ۲۸ نرون در لایه‌های پنهان پیشنهاد دادند که ورودی‌های شبکه وزن گل، عمق حفاری، خواص الاستیک سنگ، وزن روی مته، وزن گل حفاری، سایز مته، شدت جریان و خروجی شبکهٔ میزان نرخ نفوذ بود. بدیهی است که انتخاب مته‌ای با بیش‌ترین نرخ نفوذ برای ادامهٔ حفاری سبب کاهش هزینهٔ حفاری برای هر فوت می‌شود. نتایج نشان داد که تلفیق این دو الگوریتم برای انتخاب بهینه مته حفاری بسیار کارآمد است [۲۹].

مواد و روش‌ها

در علم داده‌کاوی از الگوریتم‌های مختلفی استفاده می‌شود که در سال‌های اخیر برخی از آنها نتایج بسیار قابل قبولی در دنیای واقعی از خود نشان داده‌اند. الگوریتم‌های داده‌کاوی به‌کمک نرم‌افزارهای متداول در داده‌کاوی مانند نرم‌افزار وکا[1] پیاده‌سازی شده‌اند. با توجه به گسترهٔ این الگوریتم‌ها نمی‌توان به این سوال که کدام الگوریتم بهترین الگوریتم به‌کار رفته در مورد انتخاب مته است، پاسخ دقیقی داد. علت این موضوع به هدف از تحقیق، شرایط تحقیق، تعداد داده‌ها و متغیرهای موجود و هم‌چنین بسیاری از فاکتورهای دیگر برمی‌گردد. اما بررسی و استفاده از انواع مختلف این الگوریتم‌ها سبب شده برخی از آنها به‌دلایلی از جمله سرعت و دقت کارکردشان معروف‌تر و پر استفاده‌تر شوند [۲۶]. در ادامه برخی از این الگوریتم‌ها که در این پژوهش استفاده شده به اختصار معرفی می‌شوند:

1. WEKA

۱. درخت تصمیم

درخت تصمیم مجموعه‌ای از شرط‌های منطقی است که به‌صورت یک الگوریتم با ساختار درختی برای رده‌بندی یا پیش‌بینی یک پیامد به‌کار می‌رود [۳۰]. در درخت تصمیم‌گیری، تعدادی پرسش وجود دارد و با مشخص شدن پاسخ هر پرسش، پرسشی دیگر طرح می‌شود. اگر پرسش‌ها درست و سنجیده مطرح شوند، به تعداد کمی‌از پرسش‌ها برای پیش‌بینی دسته نیاز است. عملکرد درخت تصمیم به این صورت است که یک گره ریشه در بالای آن کشیده شده و برگ‌های آن در پایین قرار دارند. هر نمونه‌ای که در گره ریشه وارد می‌شود، در یک آزمون قرار می‌گیرد تا روشن شود که این رکورد به کدام یک از گره‌های بعدی ارسال شود. بر خلاف دسته‌بندی کننده‌های تک‌مرحله‌ای رایج که هر نمونه روی تمام دسته‌ها امتحان می‌شود، در یک درخت دسته‌بندی کننده، هر قسمت از نمونه (با توجه به مشخصه‌های آن) روی زیرمجموعه‌های خاصی از دسته‌ها امتحان شده و محاسبات غیر لازم حذف می‌شوند. اگر چه الگوریتم‌های درخت تصمیم به لحاظ تنوع زیادند اما در این پژوهش تنها الگوریتم‌های LTM و $J48$ که در این تحقیق استفاده شده‌اند توضیح داده می‌شوند:

J48: الگوریتمی برای تولید درخت تصمیم الگوریتم $C4.5$ است که به‌صورت هرس و غیر هرس کار می‌کند. درخت$C4.5$ سعی می‌کند تا به‌صورت بازگشتی مجموعه داده‌ها را با استفاده از اطلاعات نرمال به زیرمجموعه‌هایی تقسیم کند که از ریشه به‌سمت برگ حرکت کرده و در نهایت به کلاس مورد نظر می‌رسد [۳۱].

LMT: درختی تصمیمی است که برگ‌ها دارای توابع رگرسیونی هستند. این الگوریتم می‌تواند با متغیرهای باینری، چندطبقه‌ای، عددی، اسمی و حتی متغیرهایی که مقادیر آن‌ها از دست رفته سروکار داشته باشد [۳۲].

۲. احتمال بیزین[1]

بیزین ساده:

یکی از روابط مهم احتمال، رابطهٔ احتمال بیز است که به‌کمک آن احتمال برچسب کلاس یک نمونه از داده‌ها تخمین زده می‌شود. استفاده از این قانون برای طبقه‌بندی، دقت و سرعت خوبی را در پایگاه داده‌های بزرگ به‌همراه دارد. در این روش فرض بر این است که

1. Bayesian

تأثیر مقدار یک صفت خاصه روی برچسب کلاس، مستقل از مقادیر دیگر صفات خاصه است و این موضوع استقلال شرطی[1] کلاس نامیده می‌شود. این فرض چنان‌که در ادامه مشاهده می‌شود، باعث ساده‌تر شدن محاسبات می‌شود.

فرض کنید C نام صفت خاصهٔ کلاسی با m مقدار متمایز در مجموعه داده‌های آموزشی $P\ (C = ci|d\ )$ باشد. برای تخمین برچسب نمونه‌ای مانند d، همهٔ احتمالات شرطی محاسبه و بیش‌ترین احتمال برچسب کلاس d را تعیین می‌کند. اگر مجموعه دادهٔ D دارای یک صفت خاصه باشد، d می‌تواند به‌صورت (٦) بیان شود [۳۳]:

$$d =< A_1 = a_1, A_2 = a_2, ..., A_n = a_n > \quad (٦)$$

و قانون بیز می‌تواند به‌صورت (۹) نشان داده شود:

$$P(C = c_i \mid A_1 = a_1, ..., A_n = a_n) = \frac{P(A_1 = a_1, ..., A_n = a_n \mid C = c_i) \times P(C = c_i)}{P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n)}$$
$$= \frac{P(A_1 = a_1, ..., A_n = a_n \mid C = c_i) \times P(C = c_i)}{\sum_{k=1}^{m} P(A_1 = a_1, ..., A_n = a_n \mid C = c_k) \times P(C = c_k)} \quad (۷)$$

مقدار $P(C = ci)$ را می‌توان از روی مجموعه داده‌های آموزشی D برحسب کسری از داده‌ها که دارای برچسب کلاس Ci هستند، محاسبه کرد. مقدار احتمال زیر نیز برای همه کلاس‌ها یک‌سان است و بنابراین تأثیری در تصمیم‌گیری ندارد [۳۳].

$$P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n) \quad (۸)$$

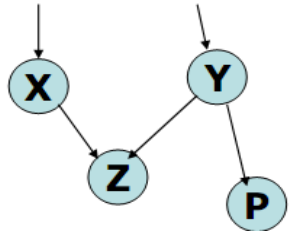
آن‌جا که می‌توان محتمل‌ترین کلاس (بزرگ‌ترین مقدار برای رابطه از میان کلاس‌های موجود) را برای نمونه‌ای از داده‌های آزمایشی تخمین زد و با توجه به این‌که مقدار مخرج رابطهٔ بیز برای همهٔ کلاس‌ها یک‌سان است، فقط کافی است صورت کسر محاسبه شود. بنابراین کافی است برای هر یک از کلاس‌های موجود، رابطهٔ (۹) محاسبه و با توجه به بیش‌ترین مقدار، کلاس نمونه‌ی آزمایشی تخمین زده شود [۳۳].

$$P(C = c_i) \times \prod_{i=1}^{n} P(A_j = a_j \mid C = ci) \quad (۹)$$

شبکه بیزی

1. Conditional Independence

شبکه‌های بیزی وابستگی‌های شرطی بین متغیرها (ویژگی‌ها) را شرح می‌دهد. با استفاده از این شبکه‌ها دانش قبلی در زمینه وابستگی بین متغیرها با داده‌های آموزش مدل طبقه‌بندی، ترکیب می‌شوند. شکل ۱ یک نمونه شبکه بیزین را نمایش می‌دهد [۳٤].



شکل ۱ .نمونه‌ای از شبکه بیزین

در شبکه بیزین، گره‌ها متغیرهایی هستند که هر کدام مجموعه مشخصی از وضعیت‌های دوبه دو ناساگاز دارند. کمان (یال) نشان‌دهندهٔ وابستگی‌های متغیرها به یک‌دیگر است. برای هر گره توزیع احتمال محلی وجود دارد که به گره وابسته است و از وضعیت والدین مستقل است [۳٤]. فرض مهم در روش بیز ساده، استقلال شرطی طبقه‌ها از یک‌دیگر است اما در عمل این وابستگی بین متغیرها وجود دارد. شبکه‌های احتمالی بیزین این نوع احتمال‌ها را بررسی می‌کند. یک شبکهٔ بیزی از دو بخش گراف غیرحلقوی و احتمال‌های شرطی تشکیل شده است اگر کمانی از Y به Z وصل شود بدین معناست که Y پدر Z است. هر کمان دانش علل و معلولی بین متغیرهای مرتبط را نشان می‌دهد. هر متغیر A با والدین $B1, B2, \ldots, Bn$ یک جدول احتمال شرطی وجود دارد در این جدول برای هر متغیر رابطهٔ آن با والدینش در نظر گرفته می‌شود [۳٤].

فرض کنید داده x با ویژگی $x1,x2,\ldots xn$ است در این صورت رابطهٔ (۱۰) احتمال شرطی با توجه به وابستگی بین متغیرها را نشان می‌دهد [۳٤].

$$p(x_1, x_2, .., x_n) = \prod_{i=1}^{n} p(x_i \mid parents(y_i)) \quad (۱۰)$$

۳. قوانین انجمنی

روش دیگری که برای طبقه‌بندی داده‌ها استفاده می‌شود، استخراج مجموعه‌ای از قوانین[1] به‌صورت شرط است. در این بخش نشان داده می‌شود که می‌توان این شروط را به‌طور

1. Rules

مستقیم از مجموعه داده‌های آموزشی استخراج کرد [۳۵].

یک قانون به‌صورت کلی(۱۱) نمایش داده می‌شود:

$$If \ \ (A_1 \ op \ v_1) \ and \ \ (A_2 \ op \ v_2) and \ ... \ Then \ Class \ = \ c_i \qquad (11)$$

که در آن $A1$ نشان‌دهندهٔ نام یک صفت خاصه، $v1$ یک مقدار مشخص، متغیر $Class$ به صفت خاصهٔ کلاس و Ci به یکی از مقادیر کلاس‌ها اشاره می‌کنند. متغیر op از میان عملگرهای مقایسه‌ای انتخاب می‌شود [۳۵].

$$Op \ = \{=, \ \ <>, <, >, <>\} \qquad (12)$$

در این بخش سمت چپ هر قانون را با واژهٔ شرط یا شروط شناخته می‌شود و هر شرط کامل (همراه با تخمین کلاس) را به‌عنوان یک قانون بیان می‌شود. چنان‌که در شکل کلی هر قانون مشاهده می‌شود، قسمت ابتدایی و سمت چپ هر قانون مجموعه‌ای از آزمون‌هایی است که روی صفات خاصه انجام شده است. میان این شروط از عملگر منطقی and استفاده می‌شود. قسمت نتیجه‌گیری قانون (سمت راست) همیشه تعیین یک کلاس از مجموعه برچسب‌های موجود در داده‌های اصلی است. چنان‌چه مجموعهٔ شروط سمت چپ یک قانون برای نمونه‌ای از داده‌ها صادق باشد، در واقع قانون مزبور، آن نمونه را پوشش می‌دهد. بنابراین می‌توان گفت یکی از اهداف ما یافتن قوانینی است که تعداد نمونه‌های بیش‌تری از مجموعه داده‌های آموزشی را پوشش دهد [۳۵]. بدین‌ترتیب مدل با تعداد قوانین کم‌تری قابل توصیف است. علاوه بر این امکان دارد نمونه یا نمونه‌هایی موجود باشند که شروط سمت چپ قانون را ارضا کنند، اما کلاسی متفاوت از کلاس تخمین زده شده به‌وسیلهٔ قانون داشته باشند. در این مورد دقت مدل نهایی کمی‌کاهش می‌یابد. در برخی از منابع می‌توان رابطه‌هایی را برای ارزشیابی یک قانون پیدا کرد که در رابطهٔ (۱۳) از رایج‌ترین آن‌ها به‌شمار می‌رود: [۳۵].

$$Coverage(rule) = \frac{N_{Cover}}{|D|} \qquad (13)$$

$$Accuracy(rule) = \frac{N_{Correct}}{N_{Cover}}$$

پس از یافتن چنین قانونی کافی است کلاس تخمین زده شده به‌وسیلهٔ این قانون به‌عنوان

کلاس نمونه‌ی آزمایشی در نظر گرفته شود [۳۵]. به‌طورکلی می‌توان گفت در همهٔ این راه حل‌ها یک نوع اولویت‌گذاری میان قوانین مطرح است. برای مثال در برخی از روش‌ها ترتیب میان قوانین اهمیت ویژه‌ای دارد. بدین‌معنی که برای یک نمونهٔ آزمایشی، باید قوانین به‌ترتیب خاصی آزمون شوند. بدین‌ترتیب اولین قانونی که نمونهٔ ازمایشی را پوشش دهد، به‌عنوان قانون اصلی برای تخمین برچسب کلاس نمونه انتخاب می‌شود. ترتیب قرارگیری این قوانین می‌تواند بر اساس اولویت کلاس‌ها تعیین شود [۳۵]. این الگوریتم‌ها پس از یافتن همهٔ قوانین مربوط به یک کلاس خاص به سراغ قوانین کلاس بعدی می‌روند. مدل‌های مختلف الگوریتم‌های آموزشی سبب شده مدل‌های مختلفی از الگوریتم‌های مبتنی بر یافت شروط ایجاد شود که از جمله معروف‌ترین آن‌ها *JRip* است [۳٦]. در الگوریتم *JRip* ابتدا کلاس‌ها در اندازه‌های بزرگ بررسی می‌شود و یک مجموعه اولیه از قوانین برای کلاس با استفاده از افزایش هرس کاهش خطا تولید می‌شود [۳٦].

٤. روش‌های مبتنی بر تشابه

همهٔ روش‌های طبقه‌بندی که تاکنون توضیح داده شد، ابتدا مدلی را با کمک داده‌ها طراحی می‌کنند و پس از آن با استفاده از مدل، برچسب کلاس نمونهٔ خواسته شده را تخمین می‌زنند. تصور کنید بدون ساختن مدل، روش با مقایسهٔ نمونهٔ آزمایشی و مجموعه داده‌ها و یافتن مشابه‌ترین نمونه قادر باشد کلاس دادهٔ آزمایشی را تخمین بزند. این روش الگوریتم‌هایی موسوم به یادگیرنده‌های تنبل است. کلمهٔ تنبل به این‌دلیل انتخاب شده است که روش تا ورود دادهٔ آزمایشی صبر می‌کند و به ساخت مدلی برای طبقه‌بندی نمی‌پردازد. این دسته از الگوریتم‌ها به تکنیک‌های کارایی برای ذخیره‌سازی و بازیابی نیاز دارند و مناسب برای پیاده‌سازی در محیط‌های موازی هستند. در ضمن این روش‌ها روی داده‌هایی که مدل پیچیده‌ای دارند نیز عملکرد خوبی دارند. چرا که راهکارهای دیگر برای ساختن مدل برای چنین داده‌هایی با مشکلاتی روب‌رو می‌شوند.

الگوریتم نزدیک‌ترین همسایگی[1]

این الگوریتم از زمانی‌که قدرت محاسباتی کامپیوترها افزایش یافت محبوب شد و یکی

1. K-nearest neighbors algorithm(KNN)

از کاربردهای رایج آن تشخیص الگوست. برای یک دادهٔ آزمایشی، الگوریتم به دنبال k نمونه از نزدیک‌ترین نمونه‌ها می‌گردد (k نمونهٔ مشابه). نزدیکی دو نمونه با به‌دست آوردن تشابه و یا فاصله‌ی میان این دو نمونه محاسبه می‌شود. فاصله و تشابه از رابطه‌های (۱۵) و (۱٦) به‌دست می‌آیند [۳۷]:

$$Dis(A_i, A_j) = Max(|1-v|, |0-v|) \quad (۱۵)$$

$$Dis(A_i, A_j) = Min(|1-v|, |0-v|) \quad (۱٦)$$

برای صفات خاصه غیرعددی کافی است حداقل یکی از آن‌ها ناقص باشد، تا فاصله برابر با یک و تشابه برابر با صفر تنظیم شود [۳۷]. روش کاربردی دیگر برای بهبود الگوریتم اولیه، محاسبهٔ فاصله و یا تشابه میان زیرمجموعه‌ای از صفات خاصه به‌جای فضای کل است. بدین‌ترتیب که در ابتدا فاصلهٔ میان نمونه‌ی آزمایشی و داده‌های ذخیره شده با توجه به زیرمجموعه‌ای از صفات خاصه به‌جای همهٔ آن‌ها محاسبه می‌شود. در این صورت چنان‌چه فاصله (تشابه) از مقدار تعریف شده‌ای بیش‌تر (کم‌تر) باشد، بدون محاسبهٔ کامل فاصله یا تشابه به سراغ داده‌ی بعدی باید رفت. به‌علاوه اگر به هر نحوی قادر باشد برخی از نمونه‌ها را از فضای جستجو خارج کند بدون شک سرعت الگوریتم افزایش می‌یابد [۳۷].

روش کا-استار[1]

یک روش طبقه‌بندی مبتنی بر تشابه است که براساس موارد آموزشی مشابه، طبقه‌بندی را انجام می‌دهد و در مقایسه با الگوریتم‌های یادگیری ماشین، نتایج مطلوبی را حاصل می‌نماید. این روش برخلاف دیگر روش‌های داده‌کاوی که براساس تابع فاصله مبنی بر آنتروپی، طبقه‌بندی را انجام می‌دهند، از تابع شباهت برای تخمین متغیرهای مختلف استفاده می‌کند. فرض اصلی طبقه‌بندی مبتنی بر تشابه، این است که موارد مشابه کلاس‌های مشابه داشته باشد [۳۸].

۵. سیستم استنتاجی نروفازی تطبیقی[2]

رایانش نرم نسل جدیدی از سیستم‌های تلفیقی هوشمند را ارائه کرده است که با تلفیق شبکه‌های عصبی و سیستم‌های فازی، تخمین جواب و بهینه‌سازی مسائل پیچیده را انجام

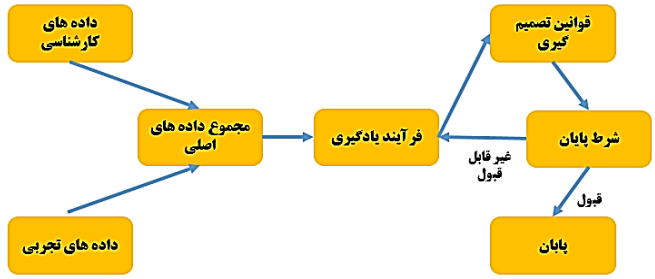
1. K-STAR
2. Adaptive neuro fuzzy inference system (ANFIS)

می‌دهند. سیستم‌های نروفازی استنتاجی تطبیقی ملقب به *ANFIS*، نمونه‌ای از این سیستم‌های تلفیقی هوشمند است که در این بخش در مورد آن بحث می‌شود [۳۹].

سیستم‌های نروفازی

برای پیاده‌سازی یک سیستم فازی به پایگاه قواعد نیاز است، حال آن‌که در بسیاری از کارهای عملی این پایگاه قواعد موجود نیست، بلکه تنها یک سری ورودی–خروجی در دسترس است که نمی‌توان رابطهٔ بین ورودی‌ها و خروجی‌های موجود را کشف کرد که دلیل آن می‌تواند پیچیدگی مسئله و یا کافی نبودن دانش شخص باشد. از این‌رو، دغدغه استخراج قانون از یک‌سری اطلاعات خام یکی از نیازهای استفاده کنندگان از سیستم‌های استنتاجی فازی است. برای این‌که بتوان این مشکل را حل کرد، شبکه‌ها عصبی مصنوعی بهترین گزینه‌ی انتخاب هستند [٤٠]. ولی مشکلی که در رابطه با شبکه‌ها عصبی مصنوعی به‌وجود می‌آید این است که این شبکه‌ها به‌عنوان یک جعبهٔ سیاه عمل می‌کنند و به ما اطلاعاتی در رابطه با دانش موجود بین ورودی‌ها و خروجی‌ها نمی‌دهند بنابراین، نحوهٔ عملکرد آن‌ها و چگونگی یافتن ارتباط میان ورودی–خروجی مشخص نیست. از این‌رو در اواخر دههٔ هشتاد ایدهٔ ترکیب شبکه‌های عصبی مصنوعی با سیستم‌های فازی مطرح شد و امروزه یکی از مباحث مورد توجه جوامع علمی است. با این ایده از یک سو هوشمندی شبکه‌ها عصبی مصنوعی در استخراج قاعده از یک‌سری داده خام به‌کمک سیستم‌های فازی می‌آید و از سوی دیگر سیستم‌های فازی به شبکه‌های عصبی مصنوعی شفافیت می‌دهند. بنابراین می‌توان فرآیند تصمیم‌گیری را به‌صورت طرحی شماتیک در شکل ۲ نشان داد و دریافت که از قسمت فرآیند یادگیری تا قسمت پایان موجود در شکل ۲ به‌وسیلهٔ روش شبکهٔ عصبی مصنوعی کنترل می‌شود. با این تفاوت که اگر از روش استنتاج فازی نیز استفاده شود می‌توان قواعد و قانون‌های ایجاد شده برای تصمیم‌گیری را نیز مشاهده کرد [٤١]–[٤٤].

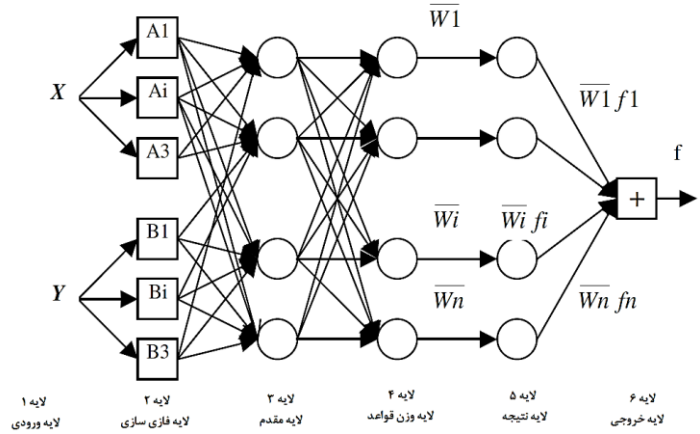
بنابراین زمانی که سیستم‌های فازی و شبکه‌های عصبی مصنوعی به‌تنهایی قادر به حل مسائل پیش روی نیستند برای حل این مشکل از تلفیق این دو سیستم فازی و عصبی استفاده می‌شود. تلفیق عصبی و فازی سیستم جدیدی را به‌وجود می‌آورد با نام *ANFIS* که به‌معنی سیستم‌های استنتاجی نروفازی تطبیقی است [٤١].



شکل ۲. فرآیند تصمیم‌گیری

ساختار سیستم‌های نروفازی

ساختار سیستم استنتاجی نروفازی تطبیقی (*ANFIS*) از ٦ لایه تشکیل شده است (شکل ۳). چنان‌که در شکل ۳ مشاهده می‌شود خروجی هر ند در لایهٔ l با عبارت Ol,i نشان داده می‌شوند که l شمارهٔ لایه و i شمارهٔ نرون لایه‌ی بعدی است.



شکل ۳. ساختار سیستم استنتاجی ANFIS [٤٥]

لایهٔ **اول**: مقادیر ورودی این لایه، داده‌های ورودی هستند که پس از خروج به مرحلهٔ فازی‌سازی منتقل می‌شوند. x و y داده‌های ورودی است [٤١].

$$O_{1,x} = x \quad (۱۷)$$

$$O_{1,y} = y \quad (۱۸)$$

لایهٔ دوم: خروجی گره‌ها در این لایه با عبارت Ol,p,i معرفی می‌شوند که p متغیرهای زبانی ورودی هستند و فرض می‌شود که هر متغیر زبانی ورودی به تعداد m تابع عضویت دارد. در

این مرحله مقادیر تابع ورودی با استفاده از فازی‌سازهای مورد نظر، فازی‌سازی می‌شوند. هر ورودی می‌تواند به تعداد لازم تابع عضویت داشته باشد. که به‌ازای هر تابع عضویت برای هر ورودی، گره تعریف می‌شود. بنابراین در لایهٔ دوم برای هر ورودی می‌توان به تعداد دلخواه گره و یا به‌عبارتی تابع عضویت داشت [۴۱].

لایه‌ی سوم:

$$O_{2,x,i} = \mu_{A_{i(x)}} \quad (۱۹)$$

$$O_{2,y,i} = \mu_{B_{i(y)}} \quad (۲۰)$$

این لایه که به لایهٔ مقدم شهرت دارد لایه‌ای است که در آن به تعداد هر گره یک قاعدهٔ فازی وجود دارد که اگر n ورودی وجود داشته باشد و هر ورودی m تابع عضویت وجود داشته باشد، به تعداد nm قاعدهٔ فازی در این لایه تشکیل می‌شود. در این لایه مقادیر وزنی هر گره تعریف می‌شود. که با عملگر $T\text{-}norm$، مقادیر وزنی مطابق رابطهٔ زیر محاسبه می‌شود. n معرف n امین قاعدهٔ فازی است [۴۱].

$$O_{3,n} = W_n = \mu_{Ai(x)} . \mu_{Bi(x)} \quad (۲۱)$$

O خروجی هر گره در این لایه حاصل تأثیر همهٔ سیگنال‌های ورودی است. در این لایه هر گره، وزن n (Wn) امین قاعدهٔ فازی را محاسبه می‌کند [۴۱].

لایهٔ چهارم: در این لایه وزن محاسبه شده در هر گره، نرمالیزه می‌شود که از رابطهٔ (۲۲) محاسبه می‌شود. در حقیقت خروجی این لایه، وزن نرمالیزه شده است.

$$O_{4,n} = \bar{W}_n = \frac{W_n}{\sum W_n} \ , \ n = 1,2,\ldots,n^m \quad (۲۲)$$

تعداد گره‌های این لایه برابر با تعداد گره‌های لایهٔ سوم یعنی nm است [۴۱].

لایهٔ پنجم: این لایه به لایهٔ نتیجه شهرت دارد. هر گره در این لایه یک گره سازگار شونده استکه به‌صورت (۲۳) تعریف می‌شود.

$$O_{5,n} = \bar{W}_n . f_n = \bar{W}_n . (p_n . x + q_n . y + r_n) \quad (۲۳)$$

Pn، qn، rn پارامترهای نتیجه هر قاعده است. این لایه هم تعداد گره‌هایی برابر با گره‌های لایه پیشین دارد [۴۱].

لایهٔ ششم: به لایهٔ خروجی ملقب است. تنها گره موجود در این لایه، عملیات دفازی کردن را با استفاده از تکنیک‌های مختلفی مانند مرکز ثقل انجام می‌دهد که رابطهٔ مرکز ثقل به‌صورت (۲٤) محاسبه می‌شود[٤١].

$$O_{6,1} = \Sigma \bar{W}_n . f_n = \frac{\Sigma W_n . f_n}{\Sigma W_n} \quad (۲٤)$$

معیارهای ارزیابی

به‌منظور کنترل دقت الگوریتم‌های داده‌کاوی لازم است تا مقادیر متغیرهای وابسته از حالت پیوسته به دودویی (موفق یا ناموفق) تبدیل شود. از این‌رو مقدار میانه (معیار گرایش به مرکز) به‌عنوان معیار تفکیک در نظر گرفته شده است. به این صورت که نمونه‌ای با بازدهی بیش‌اتر از میانه در طبقه موفق و نمونه‌هایی با بازده عملکرد کم‌تر از میانه، در طبقه ناموفق قرار داده می‌شوند. عملکرد مدل‌های دودویی (اغلب با استفاده از یک ماتریس در هم ریختگی) اندازه‌گیری می‌شود. ماتریس درهم ریختگی شامل اطلاعات ارزشمندی درباره رده‌بندی‌های واقعی و پیش‌بینی شده به‌وسیلهٔ مدل رده‌بندی است. در این مقاله، بر مبنای اطلاعات حاصل از این ماتریس، شاخص صحت کلی (AC) برای مقایسه روش‌ها در نظر گرفته شده است (جدول ۱ و جدول ۲) [٤٦].

جدول ۱. ماتریس درهم‌ریختگی عملکرد مدل پیش‌بینی بازده [٤٦]

		واقعی	
		نامناسب	مناسب
پیش‌بینی	نامناسب	منفی صحیح(TN)	مثبت کاذب(FP)
	مناسب	منفی کاذب(FN)	مثبت صحیح(TP)

جدول ۲. معرفی معیار ارزیابی [٤٦]

نام	نماد	تعریف	رابطهٔ محاسباتی
صحت	**AC**	درصد اطلاعاتی که به درستی به‌وسیلهٔ مدل طبقه‌بندی شده	$\frac{TP+TN}{TP+TN+FP+FN}$

منطقهٔ بررسی شده

موقعیت چاه‌های بررسی شده (چاه $X\text{-}1$ تا $X\text{-}7$)، در میدان نفتی باتی کوزلوکا در شهر باتمن، استان باتمن واقع در جنوب شرقی آناتولی ترکیه است. این میدان در سال ۱۹۷۸

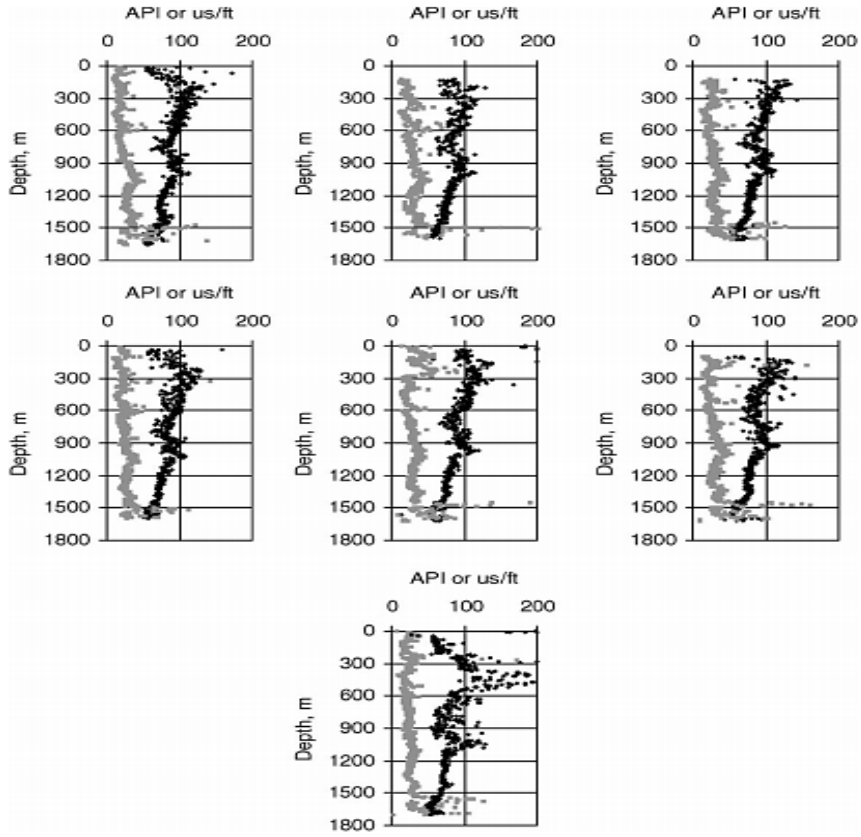
کشف شد و به‌وسیلهٔ *Petrolleri Anonim Ortakligi* توسعه یافت. تولید از این میدان نفتی در سال ۱۹۸۰ آغاز شد. مجموع ذخایر اثبات شده میدان نفتی باتی کوزلوکا حدود ۱۳۸ میلیون بشکه است و ظرفیت تولید این میدان حدود ۱۵۰۰ بشکه در روز است. در این پژوهش از داده‌های حاصل از حفاری‌های قبلی انجام شده در منطقه برای پیشنهاد مته‌های بهینه به‌منظور بهینه‌سازی و تسریع حفاری‌های آینده در منطقه (البته به‌تفکیک اعماق مختلف) استفاده شده است. چنان‌که قبلاً اشاره شد، جمع‌آوری داده‌هایی مانند داده‌های ژئومکانیکی مانند مقاومت فشاری تک‌محوری و غیره، در طول حفر چاه حتی اگر در محل امکان‌پذیر هم باشند عموماً زمان‌گیر و هزینه‌بر هستند. اما اطلاعات به‌دست آمده از چاه نمودارهای (لاگ‌های) چاه‌نگاری علاوه بر امکان‌پذیر بودن و سهولت در انجام‌شان، به‌نسبت ارزان‌تر هستند. ضمن این‌که بر اساس ماهیت‌شان به‌صورت غیرمستقیم اطلاعات لایه‌های مختلف زمین‌شناسی و مکانیک‌سنگی را نیز به‌همراه خود دارند.

در این منطقه اطلاعات ثبت شده شامل چاه نمودارهای (لاگ‌های) موج صوتی و گاما برای هر چاه است که در شکل ٤ نیز نشان داده شده است. در کنار این اطلاعات، کد مته‌های بهینهٔ تفکیک شده از بازه‌های مختلف حفاری برای هر یک از چاه‌ها نیز در دسترس بوده است و استفاده شده است (جدول ۳).

در این مقاله از اطلاعات چاه‌های *X-1* تا *X-6* به‌عنوان داده‌های آموزشی استفاده شده در مدل‌سازی به‌کار گرفته شده است و از اطلاعات چاه *X-7* به‌عنوان داده‌های آزمون برای بررسی صحت و دقت مدل‌های پیشنهادی استفاده شده و بر این اساس تعداد داده‌های آموزشی و داده‌های آزمون به‌تفکیک برای هر بازه مشخص شد (جدول ٤). توجه به پارامترهای زمین‌شناسی و ژئومکانیکی، به‌ویژه تغییر جنس لایه‌های زمین‌شناسی حین

جدول ۳. کد مته‌های بهینه برای هر اه به تفکیک عمق [٤٦]

X1		X2		X3		X4		X5		X6		X7	
Depth (m)	IADC code	Depth (m)	IADC code	Depth (m)	IADC code	Depth (m)	IADC code	Depth (m)	IADC code	Depth (m)	IADC code	Depth (m)	IADC code
150	—	150	—	150	—	150	—	150	—	150	—	150	—
403	111	302	131	680	537	480	131	479	131	235	131	401	537
544	131	681	537	910	111	695	537	650	131	347	537	918	131
715	527	874	131	1295	111	718	537	757	131	620	131	920	131
1072	111	1443	111	1457	131	1102	131	1359	131	1358	537	1156	131
1412	131					1263	111	1410	131			1400	131
						1395	131						
						1408	131						



شکل ٤. نمودار خاکستری سمت چپ مربوط به لاگ gamma و نمودار سیاه سمت راست مربوط به لاگ sonic (برای چاه‌های X-1 تاX-7) [٢٤]

پیش‌روی متهٔ حفاری به‌سمت اعماق پایین‌تر و در نظر داشتن حساسیت روش‌های هوش محاسباتی نسبت به داده‌های ورودی سبب شد که چاه‌های بررسی شده به‌تفکیک عمق و بر اساس یک حد جدایش در چند بازه تفکیک شوند. برای شناسایی حد جدایش مرزهای مختلف بازه‌های حفاری ابتدا گام پیش روی ۵۰ متر به ۵۰ متر بررسی شد. در واقع یعنی ابتدا عمق ۱۵۰ تا ۲۰۰ متری سپس ۱۵۰ تا ۲۵۰ متری و به‌همین نحو ۱۵۰ تا ۱۳۵۰ متری با پنج روش مدل‌سازی *KNN*، *ANFIS*، *Rules*، *Beays* و*DT* ، به‌صورت سعی و خطا برای یافتن بهترین نتایج برای انتخاب مناسب‌ترین بازه مدل شد. به‌عنوان مثال برای یافتن مرز جدایش اول و ایجاد بازهٔ اول از عمق ۱۵۰تا ۵۵۰ با گام مشخص شده (طول هر گام ۵۰ متر)

مدل شد. در نهایت با توجه به نتایج به‌دست آمده از انواع مدل‌سازی‌ها برای هر گام در آخر موقعیت مرز جدایش، عمق ۵۵۰ متری لحاظ شد (شکل ۵).

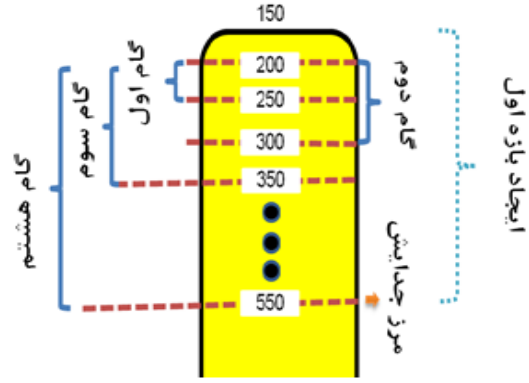
در نهایت به‌کمک روش‌های مذکور، عمق ۱۵۰ تا ۱۴۰۰ متری به چهار بازهٔ ۱۵۰ تا ۵۵۰ متری، ۵۵۰ تا ۹۷۵ متری، ۹۷۵ تا ۱۳۵۰ متری و ۱۱۵۰ تا ۱۴۰۰ متری برای مدل‌سازی نهایی تفکیک شد (جدول ۵).

جدول ٤. وضعیت داده‌های آموزشی و داده‌های آزمون استفاده شده در مدل‌سازی‌ها به‌تفکیک هر بازه

بازه	تعداد داده‌های آموزشی	تعداد داده‌های تست	درصد داده‌های تست	درصد داده‌های آموزشی
۱۵۰–۵۵۰	٦۲۵	۱۰۹	۱۵	۸۵
۵۵۰–۹۷۵	٦۸٤	۱۱٤	۱٤	۸٦
۹۷۵–۱۱۵۰	۲۸۵	٤۷	۱٤	۸٦
۱۱۵۰–۱٤۰۰	۳۹۸	٦۹	۱۵	۸۵
کل	۱۹۹۲	۳۳۹	۱۵	۸۵

جدول ۵. اطلاعات بهترین مدل‌های روش قوانین انجمنی

بازه	اطلاعات مدل برای بارگزاری در نرم‌افزار **WEKA**
۱۵۰–۵۵۰	weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
۵۵۰–۹۷۵	weka.classifiers.rules.PART -M 2 -C 0.25 -Q 1
۹۷۵–۱۱۵۰	weka.classifiers.rules.JRip -F 3 -N 1.0 -O 2 -S 1 -num-decimal-places 4
۱۱۵۰–۱٤۰۰	weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1



شکل ۵. نحوهٔ انتخاب مرز بازهٔ (حد جدایش) حفاری برای انتخاب بهینه مته

مدل‌سازی و تحلیل نتایج به‌دست آمده

عملیات مدل‌سازی و تنظیم پارامترهای مدل برای تمام پنج روشی که در بخش دوم (مواد و روش‌ها) توضیح داده شده به‌صورت مجزا برای بازه‌های عمقی مختلف انجام شد. بدین‌منظور برای هر یک از بازه‌های عمقی مختلف در نرم‌افزار ***WEKA*** و ***MATLAB*** برای هر پنج روش پارامترهای الگوریتم‌ها تنظیم شد. لازم به‌ذکر است تهیه کنندگان نرم‌افزار ***WEKA*** هنوز روش ***ANFIS*** را به برنامهٔ ***WEKA*** اضافه نکرده‌اند از این‌رو، از نرم‌افزار ***MATLAB*** برای مدل‌سازی با ***ANFIS*** استفاده شد جدول‌های ٦ تا ٩ اطلاعات مربوط به مدل‌سازی با پنج روش را نشان می‌دهند.

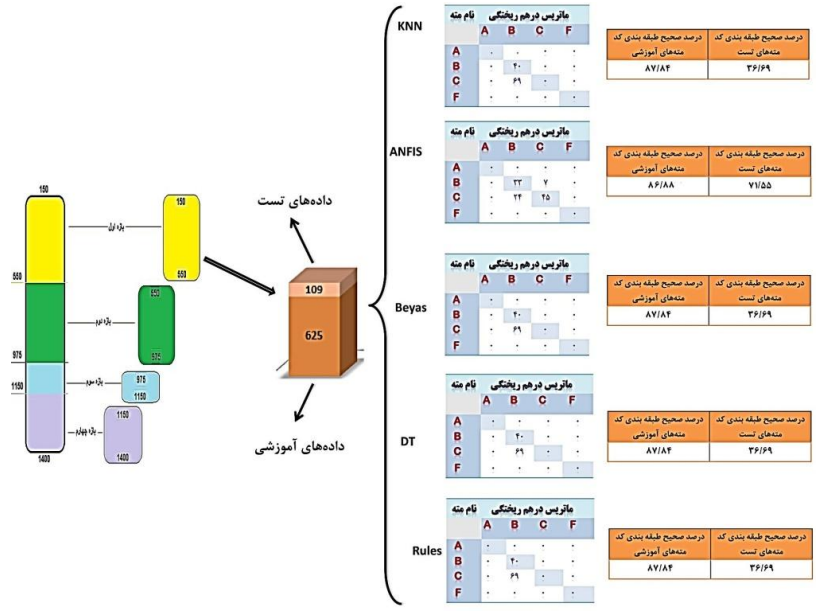
بعد از ایجاد متناسب‌ترین مدل از هر روش برای بازه‌های عمقی مختلف، مدل‌ها روی داده‌های آزمون ارزیابی انجام شد و ماتریس‌های درهم‌ریختگی آنها به‌عنوان معیاری برای بررسی کارکرد مدل‌های مختلف برای مقایسه بررسی شد. به‌صورت شماتیک در شکل‌های ٦ تا ٩ به نمایش در آمده است. در نهایت تمام نتایج مدل‌های مختلف در جدول ١٠ به‌صورت خلاصه به نمایش در آمده است.

جدول ٦. اطلاعات بهترین مدل‌های روش نزدیک‌ترین همسایه

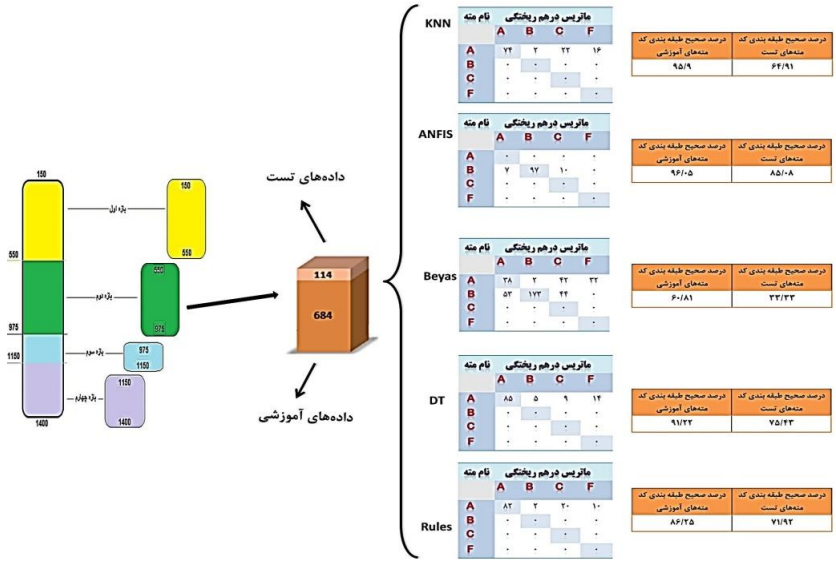
بازه	اطلاعات مدل برای بارگزاری در نرم‌افزار **WEKA**
١٥٠–٥٥٠	weka.classifiers.lazy.KStar -B 20 -M a
٥٥٠–٩٧٥	weka.classifiers.lazy.KStar -B 20 -M a
٩٧٥–١١٥٠	weka.classifiers.lazy.KStar -B 20 -M m
١١٥٠–١٤٠٠	weka.classifiers.lazy.KStar -B 20 -M a

جدول ٧. اطلاعات بهترین مدل‌های روش درخت تصمیم

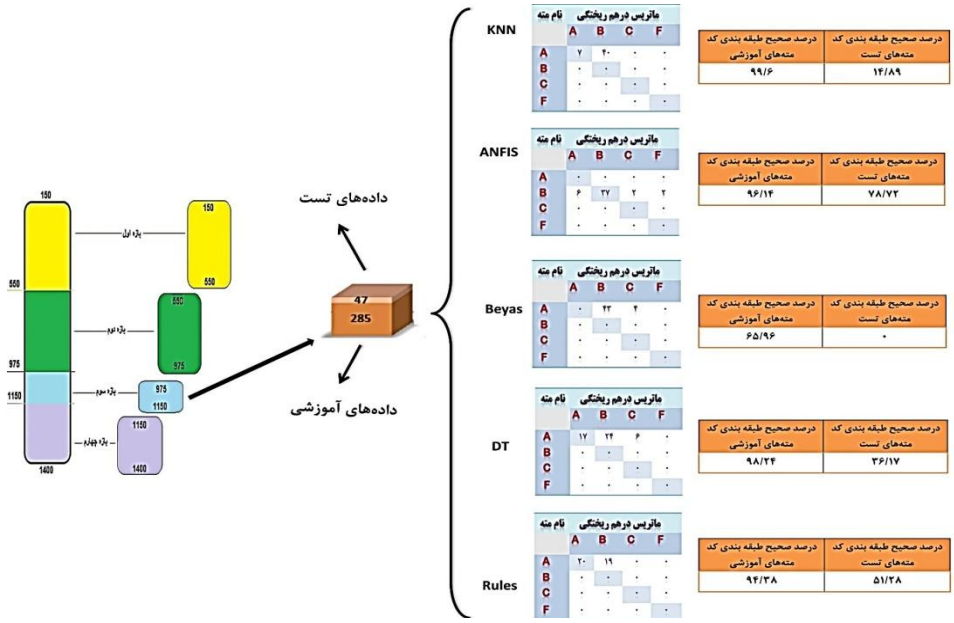
بازه	اطلاعات مدل برای بارگزاری در نرم افزار **WEKA**
١٥٠–٥٥٠	weka.classifiers.trees.LMT -I -1 -M 15 -W 0.0
٥٥٠–٩٧٥	weka.classifiers.trees.LMT -I -1 -M 15 -W 0.0
٩٧٥–١١٥٠	weka.classifiers.trees.J48 -C 0.25 -M 2
١١٥٠–١٤٠٠	weka.classifiers.trees.J48 -C 0.25 -M 2



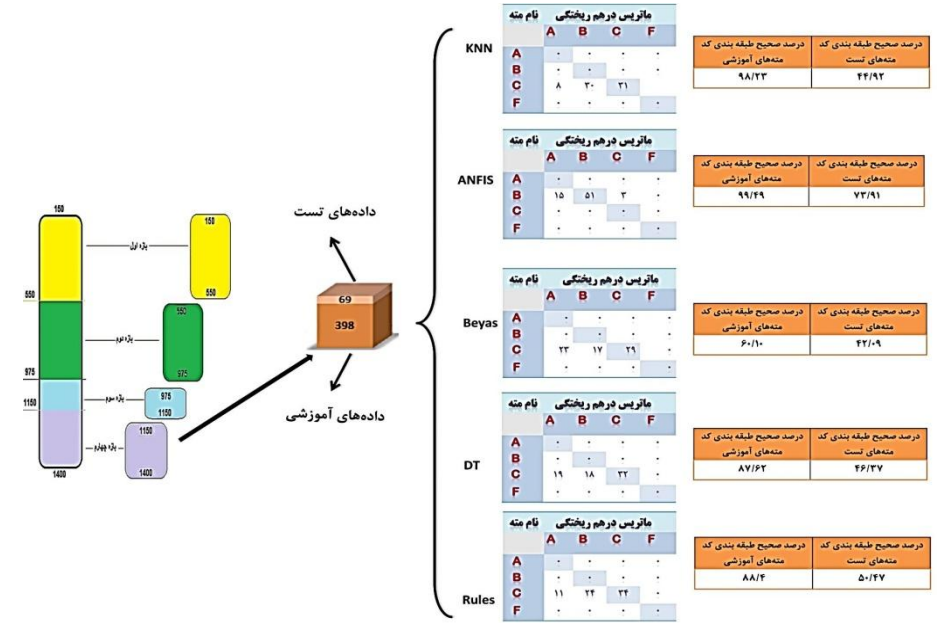
شکل ٦. نتایج روش‌های مختلف برای بازهٔ اول



شکل ۷. نتایج روش‌های مختلف برای بازهٔ دوم



شکل ۸. نتایج روش‌های مختلف برای بازهٔ سوم



شکل ۹. نتایج روش‌های مختلف برای بازهٔ چهارم

جدول ۸. اطلاعات بهترین مدل‌های روش بایزین

بازه	اطلاعات مدل برای بارگزاری در نرم‌افزار **WEKA**
۱۵۰-۵۵۰	weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.GeneticSearch -- -L 10 -A 100 -U 10 -R 1 -M -C -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
۵۵۰-۹۷۵	weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
۹۷۵-۱۱۵۰	weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
-۱۴۰۰ ۱۱۵۰	weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

جدول ۹. اطلاعات بهترین مدل‌های روش ANFIS

بازه	اطلاعات مدل برای بارگزاری در نرم‌افزار **MATLAB**				
	شعاع تأثیر	طول گام ابتدایی	نرخ کاهش طول گام‌ها	نرخ افزایش طول گام‌ها	بیشترین تعداد آغاز دوره
۱۵۰-۵۵۰	۰/۴	۰/۰۱	۰/۵	۲	۱۰۰
۵۵۰-۹۷۵	۰/۲	۰/۰۱	۰/۵	/۹	۱۰۰
۹۷۵-۱۱۵۰	۰/۳۹	۰/۰۱	۰/۵	/۹	۱۰۰
۱۱۵۰-۱۴۰۰	۰/۲	۰/۰۱	۰/۵	۱/۱	۳۰۰

جدول ۱۰. درصد صحیح طبقه‌بندی هر مدل به تفکیک بازه‌های عمقی مختلف

روش / بازه	***Rules***		***KNN***		***DT***		***Bayes***		***ANFIS***	
	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش
۱۵۰-۵۵۰	۸۷/۸۴	۳۶/۶۹	۸۷/۸۴	۳۶/۶۹	۸۷/۸۴	۳۶/۶۹	۸۷/۸۴	۳۶/۶۹	۸۶/۸۸	۷۱/۵۵
۵۵۰-۹۷۵	۸۶/۲۵	۷۱/۹۲	۹۵/۹	۶۴/۹۱	۹۱/۲۲	۷۵/۴۳	۶۰/۸۱	۳۳/۳۳	۹۶/۰۵	۸۵/۰۸
۹۷۵-۱۱۵۰	۹۴/۳۸	۵۱/۲۸	۹۹/۶	۱۴/۸۹	۹۸/۲۴	۳۶/۱۷	۶۵/۹۶	۰	۹۶/۱۴	۷۸/۷۲
۱۱۵۰-۱۴۰۰	۸۸/۴	۵۰/۴۷	۹۸/۲۳	۴۴/۹۲	۸۷/۶۲	۴۶/۳۷	۶۰/۱۰	۴۲/۰۹	۹۹/۴۹	۷۳/۹۱

از جدول ۱۰ این نتیجه حاصل می‌شود که از بین همهٔ مدل‌ها، مدل‌های *ANFIS* متهای استفاده شده در چاه آزمون (چاه شمارهٔ ۷) را در مقایسه با سایر روش‌های داده‌کاوی با دقت بیش‌تری طبقه‌بندی کرده‌اند. علت دقت زیاد روش *ANFIS* حاصل کارکرد پیچیده روش و تطابق این پیچیدگی با اطلاعات محل حفاری است. البته به‌طور جامع و دقیق نمی‌توان گفت

که پیچیدگی یک روش سبب کارکرد بهتر روش می‌شود اما می‌توان گفت در محیط‌های پیچیده بهتر است برای مدل‌سازی با روش‌های قوی‌تر شروع کرد. از این‌رو، اگر نتایج مناسب نبود از مدل‌های ساده‌تر استفاده باید کرد.

نتیجه‌گیری

نتایجی که از این تحقیق به‌دست می‌آید عبارت است از:

۱. در بررسی و مقایسه نتایج حاصل از روش‌های مختلف این مسئله اثبات می‌شود که استفاده از روش‌های هوشمند طبقه‌بندی سریع‌تر، کم هزینه‌تر و کارآمدتر از روش‌های قدیمی‌در انتخاب مته‌های حفاری بهینه هستند.

۲. روش‌های پیشین انتخاب مته نشان می‌دهد که هر یک از این روش‌ها نیازمند اندازه‌گیری پارامترهای خاص در حین حفاری و گاهاً حتی پارامترهای منحصر به‌فردی که از آزمایش‌های آزمایشگاهی و صحرایی است. بدیهی است که فراهم کردن این اطلاعات و داده‌ها عموماً سخت، زمان‌بر، و پرهزینه هستند. به‌علاوه اندازه‌گیری این پارامترها مستعد پذیرش خطاهای انسانی و دستگاهی است و جدایی از مشکلات مذکور، نیازمند شروع و انجام مراحل حفاری هستند که این خود یک مشکل بزرگ برای استفاده از این روش‌ها بوده است.

۳. روش‌های هوشمند طبقه‌بندی که در این پژوهش استفاده شده‌اند با پارامترهای مانند لاگ صوتی و لاگ گاما آموزش دیده‌اند که پارامترهای ورودی (چاه نمودارها) در هر چاهی تقریباً در دسترس هستند.

٤. بررسی عمیق‌تر نتایج حاصل از روش‌های هوشمند مختلف در کنار یک‌دیگر برای طبقه‌بندی کد مته حفاری بهینه نشان داد از بین این روش‌ها نمی‌توان گفت که کدام روش جامع‌تر و سریع‌تر است. در واقع نتایج نشان دادند پیچیدگی پردازش داده‌ها در یک روش به‌صورت مطلق نمی‌تواند معیاری برای برتری باشد. شاید با نگاه اول این حاصل شود که روش *ANFIS* نسبت به سایر روش‌ها برتری مطلق دارد، زیرا روش‌های مدل‌سازی عموماً داده محور هستند و پارامترهای مختلفی از جمله انواع داده‌ها، هم‌بستگی نوع داده‌های ورودی و پارامتر مد نظر در خروجی و غیره، می‌تواند

بر نتایج حاصل از روش‌های مختلف تأثیر محسوسی داشته باشد پس بر همین اساس تنها می‌توان گفت روش *ANFIS* برای این تیپ مدل‌سازی (ورودی لاگ گاما و لاگ صوتی- خروجی کد متهٔ بهینه حفاری) قطعاً کارآمد است. از این‌رو، در صورت وجود داده‌هایی از جنس‌های دیگر اطلاعات زمین‌شناسی باید مجدداً از تمام روش‌های مدل‌سازی برای یافتن بهترین روش مدل‌سازی استفاده شود.

۵. نتایج به‌دست آمده از دیگر روش‌ها (با صرف نظر کردن از روش *ANFIS*) نشان می‌دهد برای برخی بازه‌ها، روش‌های ساده‌تر و با پیچیدگی کم‌تر نتایج بهتری نسبت به سایر روش‌ها ایجاد کرده‌اند. مثلاً برای بازهٔ اول (۵۵۰-۱۵۰ متری) تمام چهار روش درخت تصمیم، قوانین انجمنی، احتمال بیز و مبتنی بر تشابه نتایجی کاملاً یک‌سان داشته‌اند. برای بازه‌های دیگر نیز بعد از روش *ANFIS* که بهترین عملکرد را برای مدل کردن تمام چهار بازه داشته، برای بازهٔ دوم روش درخت تصمیم و برای بازه‌های سوم و چهارم روش قوانین انجمنی در اولویت‌های بعدی هستند، پس گفتهٔ بالا مبنی بر عدم برتری روش‌ها نسبت به یک‌دیگر مجدداً تأیید شد. البته این یک ضعف برای این روش‌ها نیست زیرا محققان در استفاده از این روش‌ها به‌دنبال یافتن بهترین نتایج و درگیر فرآیند انجام شده به‌وسیلهٔ روش در پشت پرده نیستند پس اگر روشی هرچند ساده جواب مناسب و قابل قبولی ایجاد کند کاربردی و کارآمد است.

٦. از جمله نتایج این تحقیق این است که بهتر است ابتدا از روش‌های پیچیده برای شروع استفاده کرد زیرا عموماً این روش‌های پیچیده‌تر بهتر عمل می‌کنند. در غیر این صورت اگر نتایج روش‌های پیچیده مناسب نبود می‌توان استفاده از روش‌های ساده را در اولویت‌های بعدی قرار داد.

منابع

1. Bilgesu H., Al-Rashidi A., Aminian K., Ameri S., "An Unconventional Approach for Drill-Bit Selection", In: SPE Middle East Oil Show, Society of Petroleum Engineers, (2001).
2. Perrin V., Mensa-Wilmot G., Alexander W., "Drilling index-a new approach to bit performance evaluation", In: SPE/IADC drilling conference, Society of Petroleum Engineers, (1997).

3. Suba S., Christopher T., "A study on milestones of association rule mining algorithms in large databases", International Journal of Computer Applications, 47 (3) (2012) (0975-888).
4. Uboldi V., Civolani L., Zausa F., "Rock strength measurements on cuttings as input data for optimizing drill bit selection. In: SPE Annual Technical Conference and Exhibition", Society of Petroleum Engineers (1999).
5. Clegg J. M., Barton S. P., "Improved Optimisation of Bit Selection Using Mathematically Modelled Bit Performance Indices. In: IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition", Society of Petroleum Engineers (2006).
6. Thomas J., "Case History: PC Analysis of Bit Records Enhances Drilling Operations in Southern Alabama. In: SPE/IADC Drilling Conference", Society of Petroleum Engineers (1989).
7. Adam T., Millheim K., Chenevert M., Young F., "Applied drilling engineering", SPE Textbook Series, Dallas, TX 2 (1991).
8. Bourgoyne Jr. A., Millheim K., Chenevert M., Young Jr. F., "Applied Drilling Engineering", SPE Textbook Series Vol. 2, Second Printing, Society of Petroleum Engineers (1991) 152-155.
9. Teale R., "The concept of specific energy in rock drilling", In Int J Rock Mech Min Sci, Vol 1. Elsevier (1965) 57-73.
10. Mason K. L., "Three-cone bit selection with sonic logs", SPE Drilling Engineering 2 (02) (1987) 135-142
11. Dumans C. F. F., "Maidla EE PDC bit selection method through the analysis of past bit performances. In: SPE Latin America Petroleum Engineering Conference", Society of Petroleum Engineers (1990).
12. Falcao J., Maidla E., Dumans C., Dezen J., "PDC bit selection through cost prediction estimates using crossplots and sonic log data. In:

SPE/IADC Drilling Conference", Society of Petroleum Engineers (1993).

13. McAdams S. C., Gustafson W., Cooper N., "Higher Rotary Speeds Applied To Insert Bits Reduce Drilling Costs. In: SPE Annual Technical Conference and Exhibition",Society of Petroleum Engineers (1979).
14. Xu H., Tochikawa T., Hatakeyama T., "A Method for Bit Selection by Modelling ROP and Bit-Life. In: Annual Technical Meeting", Petroleum Society of Canada (1997).
15. Rabia H., Farrelly M., Barr M., "A new approach to drill bit selection. In: European Petroleum Conference", Society of Petroleum Engineers (1986).
16. Rabia H., "Oilwell drilling engineering: principles and practice", Graham & Trotman, Limited (1985).
17. Abbas R., "Using an optimum bit selection method for determined wells in Northern Iraq", Leeds University (2007).
18. Borah M. J., "Study of specific energy method of bit selection on the basis of drill bit-depth data in upper Assam oil fields", Int J Mech Eng Tech 6 (10) (2015)103-114.
19. Farrelly M., Rabia H., "Bit performance and selection: a novel approach. In: SPE/IADC Drilling Conference", Society of Petroleum Engineers (1987).
20. Javani D., Hashempour A., "The Significance of using Geological and Geomechanical Data in Selection of Optimal Bits", In: 13th ISRM International Congress of Rock Mechanics, International Society for Rock Mechanics (2015).
21. Fattahi H., "Prediction of slope stability state for circular failure: A hybrid support vector machine with harmony search algorithm", Int J Optim Civil Eng 5 (1) (2015) 103-115.
22. Fattahi H., "Prediction of earthquake induced displacements of slopes using hybrid support vector regression with particle swarm

optimization", Int J Optim Civil Eng 5 (3) (2015) 267-282.

23. Bilgesu H., Al-Rashidi A., Aminian K., Ameri S., "A new approach for drill bit selection", In SPE Eastern Regional Meeting, Society of Petroleum Engineers (2000).
24. Yılmaz S., Demircioglu C., Akin S., "Application of artificial neural networks to optimum bit selection", Comput Geosci 28 (2) (2002) 261-269.
25. Bataee M., Kamyab M., Ashena R., "Investigation of various ROP models and optimization of drilling parameters for PDC and Roller-Cone bits in Shadegan Oil field. In: International Oil and Gas Conference and Exhibition in China", Society of Petroleum Engineers (2010).
26. Bataee M., Mohseni S., "Application of Artificial Intelligent Systems in ROP Optimization: a Case Study", In SPE middle east unconventional gas conference and exhibition, Society of Petroleum Engineers (2011).
27. Edalatkhah S., Rasoul R, Hashemi A., "Bit selection optimization using artificial intelligence systems", Petrol Sci Technol 28 (18) (2010) 1946-1956.
28. Momeni M., Ridha S., Hosseini S., Liu X., Atashnezhad A., Ghaheri S., "Optimum Drill Bit Selection by Using Bit Images and Mathematical Investigation", Int J Eng Trans 30 (11) (2017) 1807-1813.
29. Momeni M., Hosseini S. J., Ridha S., Laruccia M. B., Liu X., "An optimum drill bit selection technique using artificial neural networks and genetic algorithms to increase the rate of penetration", J Eng Sci Tech 13 (2) (2018) 361-372.
30. Rokach L., Maimon O., "Decision trees. In: Data mining and knowledge discovery handbook", Springer (2005) 165-192.
31. Breiman L., Friedman J., Olshen R., Stone C., "Classification and

Regression Trees. Belmont", CA: Wadsworh. Inc (1984).

32. Chicago I. "Clementine 12 User Manual", (2007).
33. Hanson R., Stutz J, Cheeseman P., "Bayesian classification theory", (1991).
34. Friedman N., Geiger D., Goldszmidt M., "Bayesian network classifiers", Machine learning 29 (2-3) (1997) 131-163.
35. Hand D. J., "Construction and assessment of classification rules", Vol. 15. Wiley Chichester (1997).
36. Cohen W. W., "Fast effective rule induction. In: Machine Learning Proceedings", Elsevier, (1995) 115-123.
37. Wu X., Kumar V., Quinlan J. R., Ghosh J., Yang Q., Motoda H., McLachlan G. J., Ng A., Liu B., Philip S. Y., "Top 10 algorithms in data mining. Knowledge and information systems 14 (1) (2008) 1-37.
38. Cleary J. G., Trigg L. E., "K*: An instance-based learner using an entropic distance measure", In Machine Learning Proceedings, Elsevier (1995). 108-114.
39. Lin C.-T., Lee C. S. G., "Neural-network-based fuzzy logic control and decision system", IEEE T Comput 40 (12) (1991) 1320-1336.
40. Karimpouli S., Fattahi H., "Estimation of P-and S-wave impedances using Bayesian inversion and adaptive neuro-fuzzy inference system from a carbonate reservoir in Iran", Neural Comput Appl 29 (11) (2018) 1059-1072.
41. Jang J.-S., "ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man", and cybernetics 23 (3) (1993) 665-685.
42. Fattahi H.,"Indirect estimation of deformation modulus of an in situ rock mass: an ANFIS model based on grid partitioning", fuzzy c-means clustering and subtractive clustering. J Geosci, 20 (5) (2016) 681-690.

43. Fattahi H., Agah A., Soleimanpourmoghadam N., "Multi-Output Adaptive Neuro-Fuzzy Inference System for Prediction of Dissolved Metal Levels in Acid Rock Drainage: a Case Study", J AI Data Mining 6 (1) (2018) 121-132.

44. Fattahi H., "Prediction of slope stability using adaptive neuro-fuzzy inference system based on clustering methods", J Min Environ 8 (2) (2017) 163-177.

45. Drake J. T., "Communications phase synchronization using the adaptive network fuzzy inference system (anfis)", New Mexico State University (2000).

46. Kohavi R.,"Glossary of terms. Special issue on applications of machine learning and the knowledge discovery process", J Machine Learn 30 (271) (1998) 127-132.