

پیش‌بینی میزان ماسه عبوری از لوله‌ی مشبک‌کاری شده به کمک الگوریتم‌های یادگیری ماشین

محمدالیاس خدانشناس*

دانشجوی کارشناسی ارشد، مهندسی نفت، دانشکده مهندسی شیمی، نفت و گاز، دانشگاه شیراز، شیراز، ایران

آدرس پست الکترونیکی نویسنده مسئول مکاتبات: mohammadyas@hafez.shirazu.ac.ir

مقاله‌ی علمی-پژوهشی

صفحه ۱۱۸ - ۱۰۳

تاریخ پذیرش: ۱۴۰۵/۰۲/۲۰

تاریخ دریافت: ۱۴۰۵/۰۲/۲۶

چکیده

تولید ماسه از چاه‌های نفت و گاز باعث کاهش بهره‌وری و خرابی تجهیزات تولیدی می‌شود. برای کنترل این پدیده از توری‌های مهارکننده مکانیکی به‌ویژه لوله مشبک‌کاری شده استفاده می‌شود که عملکرد این توری به‌وسیله‌ی تحلیل نتایج آزمون نگهداشت ماسه تعیین می‌شود. در این پژوهش به‌منظور پیش‌بینی میزان ماسه عبوری از واحد سطح توری لوله مشبک‌کاری شده، از الگوریتم‌های یادگیری ماشین استفاده شده است. بدین منظور، نتایج حاصل از ۵۱۹ نمونه آزمایش آزمون نگهداشت ماسه‌ی ذکرشده در مطالعات پیشین به‌عنوان پایگاه داده جمع‌آوری شد که شامل: انواع ساختار و طراحی‌های مختلف توری، توزیع اندازه ذرات ماسه و شرایط آزمایشی است. پایگاه داده‌ی نهایی شامل: ۴۹۹ نمونه داده‌ی پیش‌پردازش شده است. برای ساخت مدل از الگوریتم‌های: GB، XGBoost، RF، DT، KNN، SVM، MLP و Poly Reg استفاده شد. نتایج نشان داد که الگوریتم GB با بیشترین دقت ($R^2 = 0.94$) و کمترین خطا، به‌همراه الگوریتم XGBoost بهترین عملکرد را در پیش‌بینی ارائه می‌دهند. همچنین مشخص شد که اندازه عرض شکاف‌های توری و اختلاف فشار و دبی آزمایش مهم‌ترین عوامل تأثیرگذار بر مقدار ماسه تولیدی می‌باشند. به‌علاوه، نوع طراحی توری و اندازه عرض شکاف‌های آن باید متناسب با توزیع اندازه ذرات ماسه انتخاب شود. مدل ماشین این پژوهش، مزایایی همچون: افزایش دقت پیش‌بینی نسبت به روش‌های پیشین، ارائه‌ی یک مدل تفسیرپذیر عملکرد توری و افزایش سرعت انتخاب را به همراه دارد و می‌تواند به کنترل تولید ماسه‌ی چاه‌ها کمک کند.

کلیدواژه‌ها: آزمون نگهداشت ماسه، تکمیل چاه، تولید ماسه، لوله مشبک‌کاری شده، یادگیری ماشین

۱. مقدمه

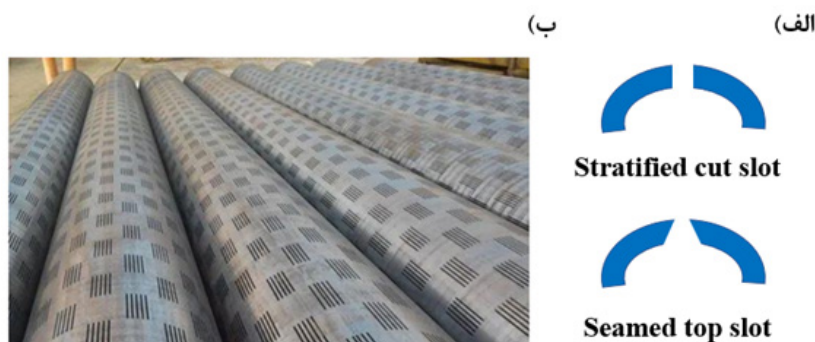
عددی، شامل: کاهش تولید، آسیب سازند و خرابی تجهیزات سطح‌الارضی و چاه را به‌همراه دارد [۱]. انتخاب یک توری مهارکننده مکانیکی^۱ مناسب از جمله: لوله مشبک‌کاری شده، می‌تواند میزان ماسه تولیدی از چاه را به حداقل برساند [۲]. لوله‌های مشبک‌کاری شده، به شکل لوله‌های دارای شکاف‌های سطحی هستند که برای جلوگیری از ورود ذرات ماسه (غالباً ذرات بزرگ‌تر ماسه) به داخل چاه طراحی شده‌اند [۳]. طراحی تعداد شکاف‌های سطح این توری با واحد SPC^۲ اندازه‌گیری می‌شود [۴، ۵]. مقدار SPC به‌گونه‌ای

2. Slot per columns

1. Mechanical Sand Screen



است. به‌طور کلی هندسه‌های: Seamed و Stratified cut دو نوع رایج طراحی شکاف هستند [۶،۷]. در (شکل ۱) تفاوت نحوه‌ی طراحی هندسه شکاف‌ها و لوله‌های مشبک‌کاری شده با تعداد شکاف‌های سطحی معین نشان داده شده است:



شکل ۱. طراحی لوله مشبک‌کاری شده: (الف) تفاوت ساختار هندسه‌های دهانه‌ی داخلی شکاف، (ب) نمونه لوله‌های مشبک‌کاری شده با تعداد معینی از شکاف‌ها [۴]

یا تکمیل چاه حفرة باز^۱ پرداخته‌اند [۲۴-۱۸]؛ بنابراین در مطالعات پیشین به کاربرد یادگیری ماشین برای افزایش دقت پیش‌بینی عملکرد لوله مشبک‌کاری شده، پرداخته نشده است [۱۹،۲۵]. در حالیکه با توجه کاربرد گسترده‌ی این توری در تکمیل چاه، تحلیل عوامل مؤثر بر عملکرد و انتخاب بهینه‌ترین طراحی توری به کمک یادگیری ماشین، می‌تواند به بهبود چشمگیری در زمینه‌ی کنترل تولید ماسه در چاه‌ها منجر شود [۶،۲۶،۲۷]. در این پژوهش با استفاده از یادگیری ماشین به پیش‌بینی ماسه عبوری از سطح توری لوله مشبک‌کاری شده پرداخته می‌شود و تأثیر عوامل مختلف بر عملکرد توری بررسی می‌شود. در نهایت، یک مدل ماشینی کارآمد برای تشخیص بهترین توری ارائه می‌شود که مقادیر ماسه عبوری را برای انواع طراحی‌های توری در یک چاه پیش‌بینی می‌کند و از طریق مقایسه‌ی نتایج، توری با کمترین تولید ماسه برای تکمیل چاه انتخاب می‌شود. در ادامه، روش انجام پژوهش شامل: جمع‌آوری داده‌ها و تعریف متغیرها، توصیف آماری و پیش‌پردازش^۹ داده‌ها،

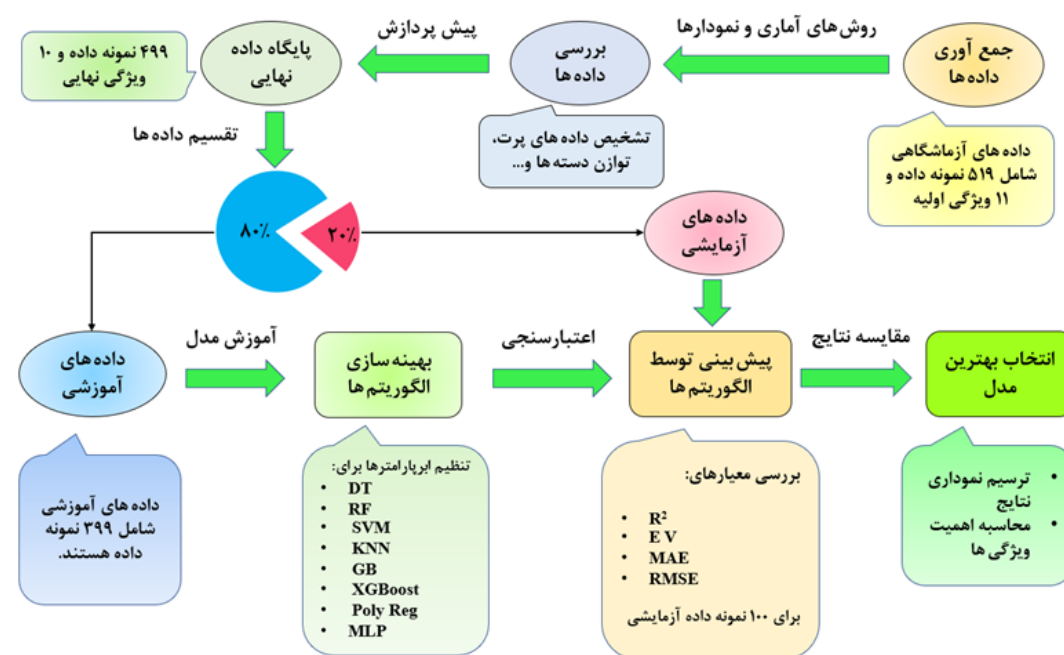
عملکرد توری انتخابی به تناسب طراحی آن با شرایط تولیدی چاه بستگی دارد [۸-۱۰]. در روش سنتی برای تشخیص بهترین توری از نتایج آزمایشگاهی (آزمون نگهداشت ماسه)^۲ استفاده می‌شود؛ به‌گونه‌ای که شرایط تولیدی چاه در آزمایش‌ها شبیه‌سازی می‌شود تا میزان ماسه عبوری از توری‌های مختلف (با طراحی متفاوت) با هم مقایسه شود [۱۱،۱۲]. انتخاب توری به روش سنتی، نیازمند صرف زمان و هزینه زیاد برای انجام آزمایش‌های متعدد است و مشکلات روش‌های آزمایشگاهی، افزایش خطای پیش‌بینی و ... را نیز به همراه دارد [۱۳]. برای رفع چالش‌های روش سنتی، استفاده از یادگیری ماشین پیشنهاد شده است که توانایی تحلیل دقیق پیچیدگی مسئله برای تعداد زیاد داده‌ها را دارد [۱۴-۱۶]. در این زمینه، عبدالرزاق و همکاران به کمک شبکه عصبی مصنوعی^۳ به پیش‌بینی میزان ماسه عبوری از توری ممتاز^۴، لوله‌ی سیم‌پیچی شده^۵ و مش فلزی^۶ پرداختند [۱۷]. سایر مطالعات مشابه نیز به پیش‌بینی میزان ماسه عبوری از توری‌های قابل انبساط^۷

1. Completion Skin
2. SRT
3. ANN
4. Premium Screen
5. Wire Wrapped
6. Metal Mesh
7. Expandable Sand Screen
8. Openhole Well Completion
9. Preprocessing

الگوریتم‌های انتخابی، روش بهینه‌سازی، روش اعتبارسنجی نتایج و انتخاب مدل نهایی شرح داده شده است.

۲. روش انجام پژوهش

به‌منظور ساخت مدل، ابتدا داده‌های آزمایشگاهی مختلف جمع‌آوری شدند و میزان ماسه‌ی عبوری از سطح توری به‌عنوان هدف^۱ پژوهش تعریف شد. سپس تأثیر هر یک از ویژگی‌ها^۲ روی تغییرات مقادیر هدف تحلیل



شکل ۲. مراحل ساخت مدل ماشینی در پژوهش حاضر

آزمایشگاهی مطالعات پیشین به‌دست‌آمده است که شامل: اطلاعات ۵۱۹ نمونه^۵ آزمایش برای انواع توری‌های لوله مشبک‌کاری شده می‌باشد [۳۴-۲۸، ۹، ۴]. برای هر نمونه، مقادیر هدف و ویژگی‌های آزمایش اندازه‌گیری و گزارش شده است. ویژگی‌های تعریف شده پژوهش، بر اساس عوامل تأثیرگذار بر تولید ماسه انتخاب‌شده‌اند که شامل: توزیع اندازه ذرات^۶ ماسه، ویژگی‌های طراحی توری و شرایط آزمایش می‌باشند. ویژگی‌های انتخابی در پژوهش در (جدول ۱) توضیح داده‌شده‌اند:

همان‌طور که در (شکل ۲) مشاهده می‌شود؛ ساخت یک مدل مناسب برای پیش‌بینی، نیازمند انجام مراحل مختلف آماده‌سازی داده‌ها برای ورود به مدل و همچنین بهینه‌سازی و اعتبارسنجی الگوریتم‌هاست. در ادامه، مراحل مربوط به آماده‌سازی داده‌ها و الگوریتم‌ها شرح داده شده است:

۲-۱. جمع‌آوری داده‌ها و تعریف ویژگی

در این پژوهش، پایگاه داده^۴ از جمع‌آوری داده‌های

1. Target
2. Features
3. Validation
4. Dataset
5. Sample
6. Particle Size Distribution



جدول ۱. ویژگی‌های موجود در پایگاه داده و معرفی آن‌ها

دسته‌بندی کلی عوامل	نماد ویژگی در پایگاه داده	واحد اندازه‌گیری	توضیحات
توزیع اندازه ذرات ماسه [۱۲]	d10	μm	۱۰ درصد دانه‌ها در نمونه ماسه از این مقدار، قطر کوچک‌تری دارند (نماینده‌ی کوچک‌ترین دانه‌ها)
	d50	μm	۵۰ درصد دانه‌ها در نمونه ماسه از این مقدار، قطر کوچک‌تری دارند (معادل مقدار میانه قطرها)
	d90	μm	۹۰ درصد دانه‌ها در نمونه ماسه از این مقدار، قطر کوچک‌تری دارند (نماینده‌ی بزرگ‌ترین دانه‌ها)
	UC	-	معیاری (ضریب یکنواختی) که میزان نزدیکی و مشابهت اندازه ذرات میانی و کوچک‌تر در نمونه به یکدیگر را نشان می‌دهد.
	SC	-	معیاری (ضریب جور شدگی) که میزان نزدیکی بین اندازه کوچک‌ترین و بزرگ‌ترین ذرات نمونه به یکدیگر را نشان می‌دهد
	Fine	-	نسبت ذرات ریز ($d < 45\mu\text{m}$) به کل ذرات برحسب درصد
SPC های ۳۰، ۴۵ و ۵۴			
ویژگی‌های طراحی توری [۳۵]	Screen type	-	تفکیک نوع هندسه طراحی شکاف (Seamed یا Stratified cut) توری بر اساس:
اندازه طول شکاف‌های توری			
	Aperture	μm	اندازه عرض شکاف‌های توری
	dp	psi	اختلاف فشار بسته‌ی ماسه ^۲
	q	ml/min	دبی جریان سیال مورد استفاده در آزمایش
شرایط آزمایش [۳۶]	fluid	-	نوع سیال آزمایشی در پایگاه داده، سیال آبی است. بنابراین یک ویژگی ثابت و بی‌تأثیر محسوب می‌شود و از پایگاه داده حذف می‌شود

۲-۲. تعریف متغیر هدف

به‌منظور مقایسه‌ی صحیح عملکرد توری‌های مختلف، رفع این چالش در پایگاه داده، مقادیر هدف به مساحت توری‌ها تقسیم شدند تا به ازای همه‌ی نمونه‌ها، واحد توری‌ها سنجیده شوند؛ زیرا تغییر مساحت توری می‌تواند بر جرم عبوری از آن تأثیر بگذارد؛ به‌منظور

1. $1\mu\text{m} = 1\text{micron} = 10^{-6}\text{ m}$

2. Sandpack

3. lb/ft²

۳-۲. توصیف آماری و پیش پردازش داده‌ها

پس از جمع‌آوری داده‌ها، انتخاب ویژگی‌ها و تعریف هدف در پایگاه داده^۱، به‌منظور پیش‌پردازش مناسب در مرحله بعدی، وجود مقادیر گم‌شده^۲ در داده‌ها بررسی شد. اطلاعات آماری نشان داد که پایگاه داده فاقد مقادیر گم‌شده است. همچنین محدوده‌ی تغییرات داده‌های هر

یک از متغیرها به‌وسیله‌ی نمودارهای گرافیکی و روش‌های آماری بررسی شد. روش‌های آماری به محاسبه‌ی مقادیر میانگین، انحراف معیار^۳، کمینه^۴، میانه^۵ و بیشینه^۶ برای هر یک از متغیرها می‌پردازند. اطلاعات آماری و توصیف داده‌های پایگاه داده، در (جدول ۲) ذکر شده است:

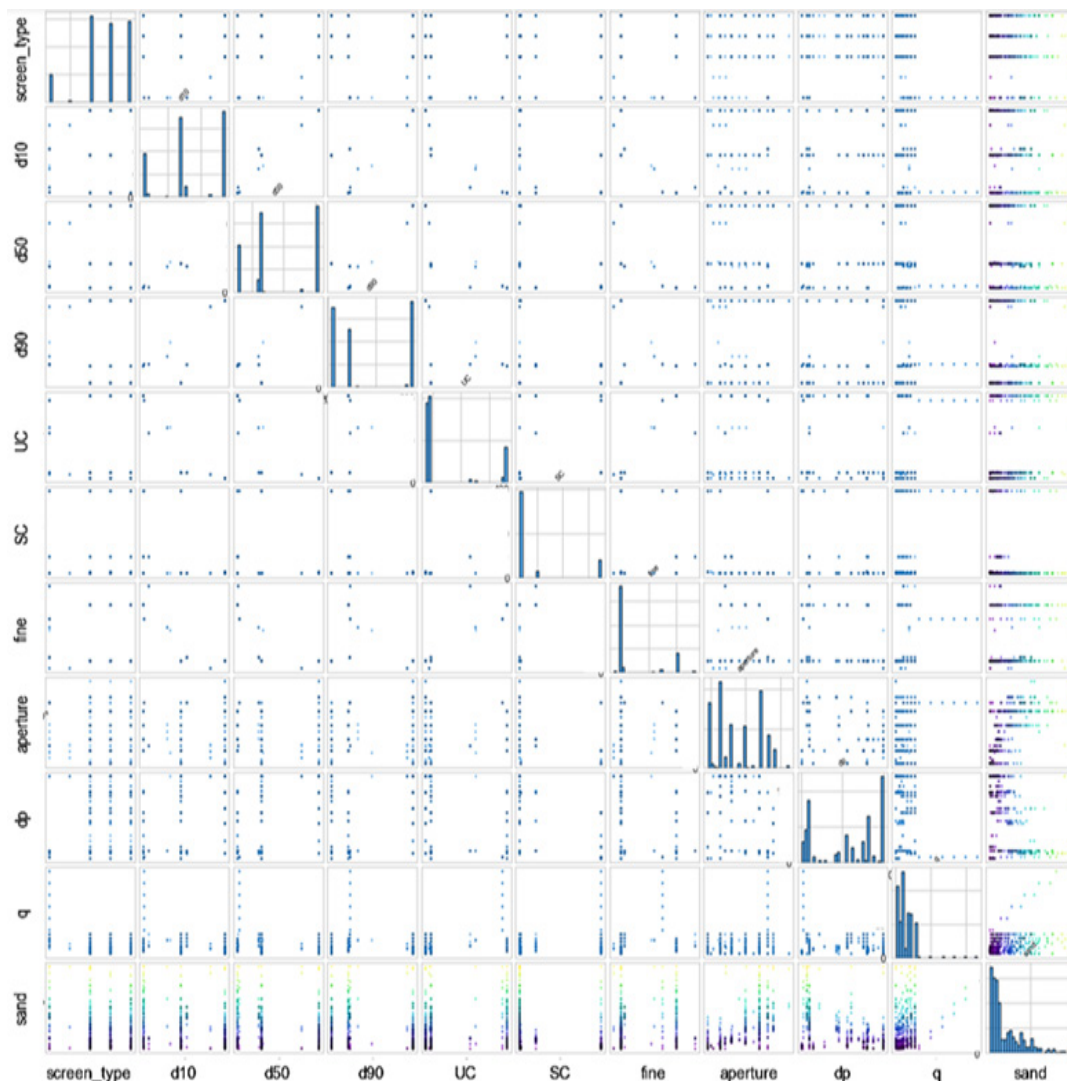
جدول ۲. اطلاعات آماری داده‌های هر یک از متغیرها

معیار	ستون	d10	d50	d90	UC	SC	fine	aperture	dp	q	sand
میانگین		۸۴/۲۵	۱۹۲/۰۸	۲۸۳/۶۵	۳/۵۴	۸/۷۶	۷/۱۱	۴۴۱/۰۶	۵۴۷/۱۰	۱۳۵/۴۹	۰/۰۰۲۸
انحراف معیار		۴۵/۸۹	۵۲/۲۸	۴۴/۷۶	۷۵/۲	۶۵۳۸/۱۱	۴/۰۷	۱۵۸/۴۵	۸۷/۳۷۶	۰۷/۱۳۶	۰/۰۰۳۰
کمینه		۱۴	۱۲۷	۲۰۸	۱/۸۸	۳/۲۰۰۳	۶۸/۳	۲۳۰	۷/۱۸	۲/۱	۰
میانه		۷۰	۱۶۵	۲۴۳	۲/۳۶	۳/۲۰۰۳	۵	۴۰۰	۶۱۰	۳۳/۱۱۰	۰/۰۰۱۴
بیشینه		۱۳۶	۲۵۶	۳۷۸	۹/۰۷	۳۵	۳۹/۱۸	۶۶۰	۱۰۰۰	۴۸/۸۶۷	۰/۰۱۳۰

با بررسی نتایج حاصل از (جدول ۲) مشخص شد که بازه‌ی تغییرات مقادیر متغیرهای پایگاه داده با یکدیگر متفاوت‌اند. این پدیده باعث سوگیری^۷ پیش‌بینی مدل به سمت متغیرهایی می‌شود که دارای بازه‌ی بزرگ‌تری هستند. به‌منظور رفع این مشکل، داده‌های هر یک از متغیرها به بازه‌ی (۰، ۱) نرمال‌سازی^۸ شدند [۳۷]. داده‌های متنی^۹ متغیر Screen_type نیز به کمک روش Encoder

به داده‌های عددی تبدیل شدند تا برای فرآیند آموزش الگوریتم‌ها آماده شوند. بهترین روش تشخیص داده‌های پرت^{۱۰} با توجه به پایگاه داده، بررسی و تحلیل وضعیت تجمع و مشابهت داده‌ها به کمک نمودارهای توزیع^{۱۱} و پراکندگی^{۱۱} است. در (شکل ۳) نمودارهای توزیع نرمال هر یک از متغیرها و نمودارهای پراکندگی متغیرها نسبت به یکدیگر نشان داده شده است:

1. Missing Values
2. Standard Deviation
3. Minimum
4. Median
5. Maximum
6. Bias
7. Normalization
8. Categorical
9. Outliers
10. Distribution Plot
11. Scatter Plot



شکل ۳. نمودارهای توزیع نرمال و پراکندگی متغیرها

نمونه داده‌ی پیش‌پردازش شده است.

۲-۴. تقسیم داده‌های آزمایشی و آموزشی

در این مرحله، داده‌های پیش‌پردازش شده به دو بخش: داده‌های آموزشی و داده‌های آزمایشی تقسیم شدند. به‌طوری‌که ۸۰ درصد داده‌ها به آموزش الگوریتم‌ها و ۲۰ درصد داده‌ها به ارزیابی نتایج پیش‌بینی‌شده توسط الگوریتم‌ها اختصاص داده شدند. مبنای انتخاب داده‌های آموزشی و آزمایشی از پایگاه داده به‌وسیله‌ی اعتبارسنجی متقابل^۱ (با قراردادن $CV=5$) انجام شد تا همه‌ی داده‌ها در فرآیند آموزش و آزمایش مدل شرکت داده شوند و نتایج مدل دچار سوگیری نشود.

با توجه به (شکل ۳) مشخص شد که توزیع داده‌های دبی جریان سیال دارای ۲۰ نمونه داده‌ی پرت است که این داده‌ها از پایگاه داده حذف شد. همچنین با بررسی نقاط دارای بیشترین فاصله نسبت به مرکز داده‌ها (نقاط زردرنگ) در نمودارها مشخص شد که این نقاط بر اساس نتایج آزمایشات تکرارپذیر معتبر به‌دست‌آمده‌اند و داده‌ی پرت محسوب نمی‌شوند. با توجه به نمودار توزیع نرمال برای ویژگی نوع توری مشخص شد که تعداد نمونه‌های انواع توری متفاوت است که ممکن است باعث سوگیری مدل شود؛ بنابراین برای رفع نامتوازن بودن تعداد نمونه در دسته‌های نوع توری از روش Smooth استفاده شده است. پس از تکمیل پیش‌پردازش، پایگاه داده‌ی نهایی برای آموزش الگوریتم‌ها شامل: ۴۹۹

1. Cross Validation

۵-۲. الگوریتم‌های مورد استفاده و روش بهینه‌سازی

به‌منظور ساخت مدل ماشینی از هشت الگوریتم، شامل: DT، RF، SVM، KNN، GB، XGBoost، رگرسیون از نوع Poly Reg و یک شبکه عصبی MLP استفاده شده است. در ادامه به بهینه‌یابی^۱ ابرپارامترها^۲ در الگوریتم‌های انتخابی پرداخته شد و هر یک از الگوریتم‌ها با تنظیم مقادیر مختلف ابرپارامترهایشان به روش جست‌وجوی شبکه‌ای^۳ روی پایگاه داده آموزش دیدند که بر اساس مقایسه‌ی دقت پیش‌بینی‌ها به‌ازای حالت‌های مختلف، بهترین مقادیر ابرپارامترهای هر الگوریتم مشخص شد. برای الگوریتم‌های مبتنی بر درخت (DT، RF، GB، XGBoost) ابرپارامترهای: حداکثر عمق^۴، تعداد درخت^۵، حداقل نمونه برای تقسیم گره^۶ و ... بهینه‌سازی شدند. ابرپارامتر نرخ یادگیری^۷ نیز برای الگوریتم‌های MLP، GB، XGBoost بهینه شد. برای شبکه عصبی حالت‌های مختلف تعداد لایه‌های پنهان^۸، تعداد نورون‌ها، حداکثر تکرارهای آموزش^۹، توابع فعال‌سازی^{۱۰} و ... آزمایش شدند. در الگوریتم SVM ابرپارامترهای: Kernel، C، Gamma و Epsilon بهینه شدند. برای الگوریتم KNN ابرپارامترهای: تعداد همسایگی^{۱۱}، وزن‌ها^{۱۲} و معیار فاصله^{۱۳} بررسی و بهینه شدند. برای رگرسیون نیز ابرپارامتر α بهینه شد و یک تحلیل حساسیت جداگانه برای بهینه‌سازی درجه‌ی چندجمله‌ای‌ها انجام شد و معادلاتی با حداکثر درجات ۱ تا ۵ ساخته شدند که بر اساس مقایسه‌ی دقت پیش‌بینی‌های این معادلات، یک معادله‌ی درجه ۲ به‌عنوان بهینه‌ترین معادله انتخاب شد.

۶-۲. تعیین تأثیر ویژگی‌ها روی هدف

به‌منظور شناسایی و تحلیل اثر هر یک از ویژگی‌ها بر روی تغییرات هدف در پایگاه داده‌ی موجود، از ماتریس ضرایب همبستگی پیرسون^{۱۴} استفاده شده است که شیب تغییرات مقادیر هر متغیر را نسبت به سایر متغیرها محاسبه می‌کند. در ماتریس ضرایب همبستگی، ضرایب مثبت نشان‌دهنده‌ی رابطه مستقیم بین دو متغیر و ضرایب منفی نیز رابطه معکوس را مشخص می‌کنند. با بررسی مشخص می‌شود که هر یک از ویژگی‌ها چه تأثیری بر ماسه‌ی عبوری از توری دارند که این نتایج در بهینه‌سازی تکمیل چاه مؤثر است.

۷-۲. اعتبارسنجی نتایج الگوریتم‌ها و نحوه انتخاب بهترین مدل

به‌منظور اعتبارسنجی دقت پیش‌بینی الگوریتم‌ها، مقادیر هدف برای داده‌های آموزشی و آزمایشی توسط الگوریتم‌ها پیش‌بینی شدند. برای مقایسه‌ی بین مقادیر پیش‌بینی‌شده و مقادیر واقعی هدف از معیارهای دقت: R^2 و EV و معیارهای سنجش خطای: RMSE و MAE استفاده شده است. معیارهای R^2 و EV میزان مطابقت پیش‌بینی‌ها و واقعیت را نشان می‌دهند. در این دو معیار، مقدار ۱ نشان‌دهنده‌ی مطابقت کامل پیش‌بینی‌ها با مقادیر واقعی هدف است و مقادیر کوچک‌تر از ۱، نشان‌دهنده‌ی دقت کمتر پیش‌بینی می‌باشند. معیارهای دقت انتخابی به‌کاررفته در پژوهش در ادامه توضیح داده‌شده‌اند:

R^2 : معیاری که میزان تغییرپذیری مقادیر پیش‌بینی‌شده نسبت به تغییر مقادیر واقعی هدف را محاسبه می‌کند و عموماً در پایگاه داده‌ی دارای مقادیر

1. Optimization
2. Hyper parameters
3. Grid Search
4. Max_depth
5. N_estimators
6. Min_samples_split
7. Learning_rate
8. Hidden layers
9. Max_iter
10. Activation functions
11. N_neighbors
12. Weights
13. Metric
14. Pearson Correlation Coefficients Matrix

هدف پیوسته^۱ به عنوان معیار اصلی ارزیابی دقت مدل توصیه می شود [۳۸]. نحوه ی محاسبه ی R^2 در رابطه (۱) نشان داده شده است:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

که در رابطه (۱) n تعداد نمونه ها، y_i مقدار واقعی، $\hat{y}$ مقدار پیش بینی شده و $\bar{y}$ مقدار میانگین مقادیر واقعی نمونه های هدف را نشان می دهد.

EV: این معیار نشان می دهد که پیش بینی های مدل، چه میزان از واریانس مقادیر هدف را توضیح می دهند. این معیار نسبت به R^2 حساسیت کمتری به مقادیر پیش بینی پرت نشان می دهد و به عنوان یک معیار ارزیابی مناسب برای داده های آزمایشگاهی قابل استفاده است. از مقایسه ی مقدار اختلاف بین معیارهای R^2 و EV میزان سوگیری مدل ها مشخص می شود. نحوه ی محاسبه ی دقت مدل به روش EV در رابطه (۲) نشان داده شده است:

$$EV = \frac{\sum_{i=1}^n (y_i - \hat{y} - (\bar{y} - \bar{\hat{y}}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

که در رابطه (۲) n تعداد نمونه ها، y_i مقدار واقعی، $\hat{y}$ مقدار پیش بینی شده و $\bar{y}$ میانگین مقادیر واقعی، $\bar{\hat{y}}$ میانگین مقادیر پیش بینی شده ی نمونه ها را نشان می دهد [۳۹]. برای ارزیابی دقیق تر عملکرد الگوریتم ها، معیارهای خطای RMSE و MAE نیز به کار گرفته شده اند که می توانند اختلاف پیش بینی ها را ارزیابی کنند. معیارهای خطای انتخابی به کاررفته در پژوهش در ادامه توضیح داده شده اند:

MAE: معیاری که میانگین قدر مطلق اختلاف پیش بینی ها و مقادیر واقعی را محاسبه می کند و به عنوان معیار خطای اصلی برای ارزیابی عملکرد مدل در این پژوهش در نظر گرفته شده است؛ زیرا یک سنجش قابل درک از مقدار خطای مورد انتظار در پیش بینی مقادیر هدف ارائه می دهد که با همان واحد اندازه گیری مقادیر هدف بیان می شود. به علاوه برای داده های آزمایشگاهی که ممکن است دارای خطای ذاتی یا اندازه گیری باشند؛ مناسب تر است. نحوه ی محاسبه ی MAE در رابطه (۳) نشان داده شده است:

$$MSE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| \quad (3)$$

که در رابطه (۳) n تعداد نمونه ها، y_i مقدار واقعی، $\hat{y}$ مقدار پیش بینی شده ی نمونه هاست.

RMSE: این معیار به محاسبه ی ریشه ی میانگین مربعات اختلاف مقادیر پیش بینی شده و مقادیر واقعی می پردازد و به عنوان یک معیار کمکی در کنار MAE به کار گرفته شده است که روی پیش بینی های پرت مدل حساسیت بیشتری نشان می دهد. براساس این معیار، خطاهای بزرگ در پیش بینی جریمه بیشتری می شوند و روی نمونه هایی که مدل در یادگیری آن ها عملکرد ضعیف تری داشته است؛ تمرکز بیشتری دارد تا تأثیر آن ها را روی مدل ارزیابی کند که این مزیت برای کاربرد صنعتی پژوهش مناسب است. نحوه ی محاسبه ی RMSE در رابطه (۴) نشان داده شده است:

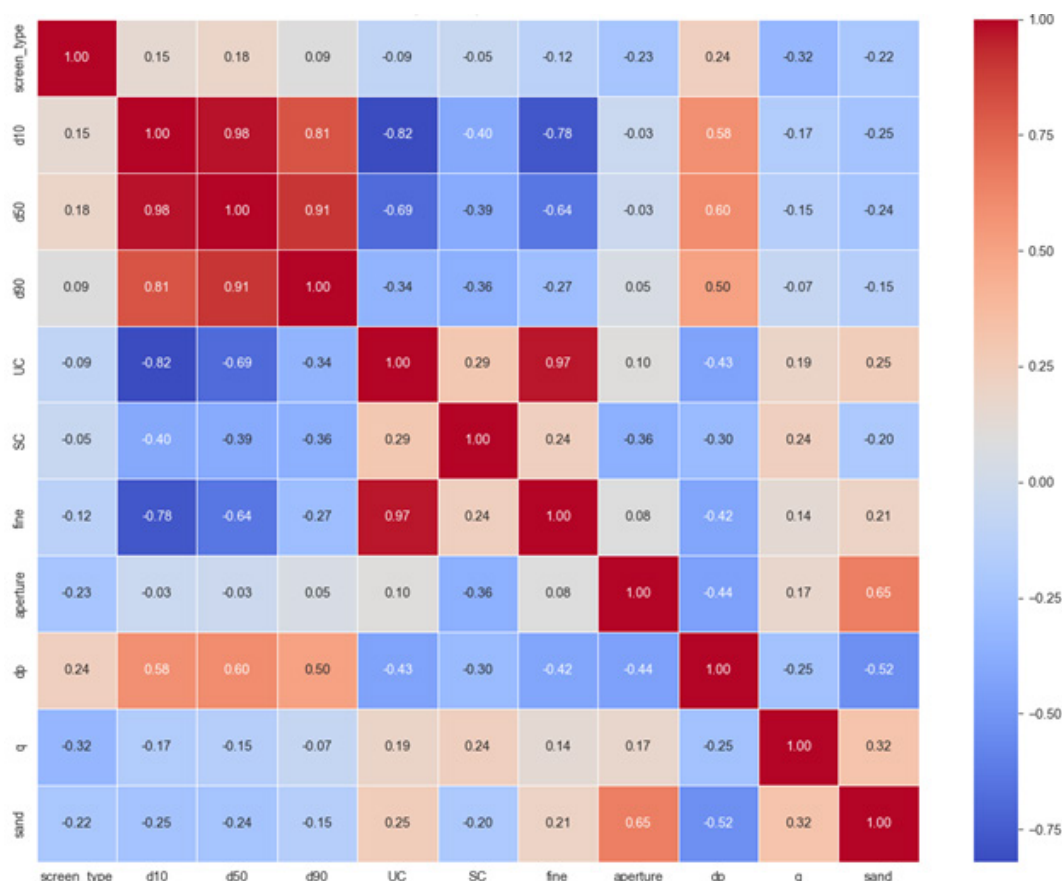
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (4)$$

که در رابطه (۴) n تعداد نمونه ها، y_i مقدار واقعی، $\hat{y}$ مقدار پیش بینی شده ی نمونه هاست [۴۰].

در مرحله ی بعد، به مقایسه ی عملکرد الگوریتم های مختلف به وسیله ی معیارهای دقت و خطا پرداخته شد و برای انتخاب مدل برتر بین الگوریتم های دارای عملکرد مشابه، از رسم نمودارهای ارزیابی استفاده شد که مقادیر پیش بینی شده را بر حسب مقادیر واقعی هدف نشان می دهند. در این نمودارها میزان مطابقت نتایج پیش بینی با مقادیر واقعی (تجمع نقاط حول نیمساز) قابل مشاهده است که به منظور تشخیص بهترین مدل استفاده شده است.

۳. نتایج مدل ها

به منظور بهره گیری از روش های یادگیری ماشین روی پایگاه داده، همبستگی هر یک از ویژگی ها با هدف پژوهش بررسی شد. تحلیل رابطه ی بین متغیرهای مختلف به کمک ماتریس ضرایب همبستگی امکان پذیر است؛ بنابراین به وسیله ی این روش، می توان به اولویت بندی و تحلیل عوامل مؤثر بر تولید ماسه پرداخت. در (شکل ۴) ماتریس ضرایب همبستگی نشان داده شده است:



شکل ۴. ماتریس ضرایب همبستگی محاسبه شده در پایگاه داده

می‌دهد در شرایطی که قطر ذرات مختلف نمونه ماسه به یکدیگر نزدیک‌تر باشد؛ کنترل تولید ماسه بهتر صورت می‌گیرد. در نهایت می‌توان دریافت که ویژگی‌های طراحی توری (اندازه عرض شکاف‌های توری (۰/۶۵) و نوع توری (۰/۲۲)) تأثیر قابل‌توجهی بر میزان ماسه عبوری از توری دارند و باید متناسب با توزیع اندازه ذرات ماسه و خصوصیات چاه تولیدی انتخاب شوند.

در ادامه‌ی پژوهش به فرآیند آموزش و بهینه‌سازی الگوریتم‌های انتخابی بر روی داده‌های آموزشی پرداخته شد. پس از تکمیل فرآیند آموزش برای ارزیابی دقت پیش‌بینی الگوریتم‌ها، مقادیر هدف برای داده‌های آزمایشی پیش‌بینی شدند. برای انتخاب مدل برتر از روش‌های اعتبارسنجی استفاده شد و الگوریتم‌ها براساس نتایج به‌دست‌آمده از اعتبارسنجی، به ترتیب اولویت معیارها رتبه‌بندی شدند. در (جدول ۳) خلاصه‌ی نتایج اعتبارسنجی هر یک از الگوریتم‌ها روی داده‌های آزمایشی نشان داده شده است:

با توجه به نتایج حاصل از ماتریس ضرایب همبستگی (شکل ۴)، مشخص شد که اندازه عرض شکاف‌های توری (مقدار ضریب ۰/۶۵ با هدف) بیشترین تأثیر مستقیم را بر مقدار ماسه عبوری از توری دارد. اختلاف فشار بسته‌ی ماسه (۰/۵۲-) بیشترین تأثیر معکوس را بر مقدار ماسه عبوری از توری دارد. دبی سیال (۰/۳۲) نیز تأثیر مستقیم و قابل‌توجهی در تولید ماسه دارد. همچنین در بررسی ویژگی‌های مرتبط با توزیع اندازه ذرات ماسه، مشخص شد که با افزایش Fine میزان ماسه عبوری افزایش می‌یابد. افزایش اندازه قطر دانه‌ها، باعث می‌شود که ذرات ماسه کمتری توانایی عبور از شکاف‌های توری را داشته باشند؛ بنابراین می‌توان نتیجه گرفت که اگر توزیع اندازه ذرات ماسه‌ی چاه دارای دانه‌بندی کوچک‌تری باشد؛ مقدار تولید ماسه‌ی توری بیشتر می‌شود. رابطه‌ی UC با میزان ماسه عبوری نشان می‌دهد که مشابهت اندازه قطر دانه‌های متوسط و دانه‌های کوچک‌تر باعث افزایش تولید ماسه از توری می‌شود. بررسی SC نشان

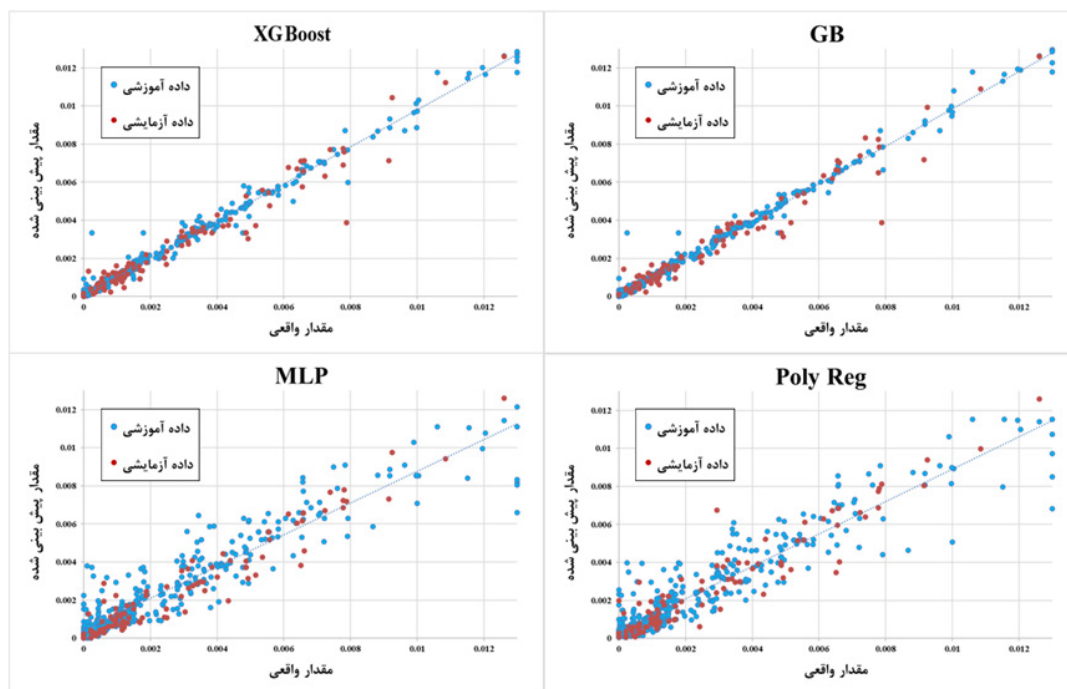


جدول ۳. اعتبارسنجی الگوریتم‌ها روی داده‌های آزمایشی به کمک معیارهای دقت و خطا

Model	R2	EV	MAE ($\times 10^{-4}$)	RMSE ($\times 10^{-4}$)
XGBoost	۰/۹۴	۰/۹۴	۴/۱۵	۷/۵۹
GB	۰/۹۴	۰/۹۴	۴/۳۷	۸/۴۴
MLP	۰/۸۸	۰/۸۹	۶/۷۶	۱۰/۷۱
Poly Reg	۰/۸۸	۰/۸۹	۶/۳۴	۹/۵۹
RF	۰/۸۷	۰/۸۸	۶/۷۱	۱۱/۶۹
DT	۰/۷۹	۰/۸۰	۸/۰۲	۱۲/۵۷
SVM	۰/۷۵	۰/۷۶	۷/۵۷	۱۳/۴۶
KNN	۰/۷۱	۰/۷۲	۷/۸۹	۱۴/۵۷

با توجه به (شکل ۵) با توجه تمرکز نقاط روی نیمساز نمودارها مشخص شد که دقت الگوریتم‌های XGBoost و GB در پیش‌بینی داده‌های آموزشی و آزمایشی بسیار به یکدیگر نزدیک است و هر دو الگوریتم روی داده‌های آموزشی و آزمایشی عملکرد بسیار مناسبی را ارائه دادند. همچنین با توجه به وضعیت انطباق داده‌ها با مقادیر پیش‌بینی شده، عملکرد مدل GB در مقایسه با XGBoost اندکی بهتر است. به توجه به نمودارها مشاهده می‌شود که این دو الگوریتم در پیش‌بینی مقادیر ماسه‌ی عبوری کمتر (توری‌های بهینه)، عملکرد بسیار دقیقی ارائه می‌دهند و به ازای مقادیر ماسه‌ی عبوری بیشتر (توری‌های متوسط یا ضعیف) اختلاف محسوس‌تری در پیش‌بینی وجود دارد. از آنجاکه برای انتخاب گزینه‌های تکمیل چاه، تمرکز بیشتر روی توری‌های بهینه است؛ به‌کارگیری این مدل‌ها برای انتخاب توری بسیار کارآمد و دقیق است. با مشاهده نمودارهای دارای نقاط پراکنده‌ی مربوط به الگوریتم‌های MLP و Poly Reg مشخص می‌شود که این دو الگوریتم نسبت به الگوریتم‌های برتر، دقت کمتری در پیش‌بینی داده‌ها ارائه می‌دهند؛ به‌ویژه در پیش‌بینی مقادیر بزرگ‌تر هدف، اختلاف پیش‌بینی‌ها افزایش یافته است. به‌علاوه با توجه به وضعیت داده‌های آزمایشی در این دو الگوریتم، مشخص می‌شود که اکثر نمونه‌ها، کمتر از مقدار واقعی پیش‌بینی‌شده‌اند. این مشکل به‌کارگیری این دو الگوریتم را در مهار ماسه‌ی چاه نامناسب می‌کند.

با توجه به (جدول ۳) می‌توان دریافت که الگوریتم‌های XGBoost و GB نسبت به سایر الگوریتم‌ها دقت بیشتری ارائه می‌دهند. این دو الگوریتم با دقت ۰/۹۴ و مقادیر خطای مشابه در پیش‌بینی هدف، عملکرد بسیار نزدیکی به یکدیگر دارند. همچنین الگوریتم‌های MLP و Poly Reg نیز عملکرد مشابهی در پیش‌بینی نشان می‌دهند. اختلاف مقادیر معیارهای R^2 و EV برای همه‌ی الگوریتم‌ها کمتر از ۰/۰۲ است که یادگیری بدون سوگیری الگوریتم‌ها روی داده‌ها را نشان می‌دهد. بر اساس معیار MAE مشخص می‌شود که برای کاربرد صنعتی، الگوریتم‌های XGBoost و GB نسبت به سایر الگوریتم‌ها مناسب‌تر هستند؛ زیرا مقادیر پیش‌بینی‌شده و مقادیر واقعی همخوانی بسیار دقیقی با یکدیگر دارند و میانگین اختلاف این مقادیر تقریباً صفر است. سایر الگوریتم‌ها نیز محدوده‌ی قابل قبولی از خطا را ارائه می‌دهند. بررسی معیار RMSE نشان می‌دهد که الگوریتم‌های XGBoost و GB نسبت به سایر الگوریتم‌ها مقادیر پرت کمتری در پیش‌بینی ارائه می‌دهند. با توجه به مشابهت نتایج اعتبارسنجی بین الگوریتم‌ها، به‌منظور مقایسه‌ی بهتر عملکردها و انتخاب بهترین مدل ماشینی، نمودارهای پیش‌بینی چهار الگوریتم برتر روی پایگاه داده در (شکل ۵) نشان داده‌شده است. در نمودارها، نقاط آبی‌رنگ نشان‌دهنده‌ی داده‌های آموزشی و نقاط قرمز رنگ نشان‌دهنده‌ی داده‌های آزمایشی هستند.



شکل ۵. نتایج پیش‌بینی الگوریتم‌ها روی داده‌های آموزشی و آزمایشی

۴. نتیجه‌گیری

در این پژوهش با استفاده از داده‌های آزمون نگهداشت ماسه، به پیش‌بینی مقدار ماسه عبوری در واحد سطح برای توری‌های لوله مشبک‌کاری شده پرداخته شد. برای ساخت پایگاه داده از نتایج ۵۱۹ نمونه آزمایش موجود در مطالعات پیشین استفاده شد و حداکثر تعداد عوامل مؤثر بر تولید ماسه در آزمایش‌ها در نظر گرفته شد. به‌منظور ساخت مدل ماشینی نیز از الگوریتم‌های KNN، SVM، RF، DT، XGBoost، GB، Poly Reg و MLP استفاده شد. نتایج نشان داد که:

۱. الگوریتم‌های مبتنی بر تقویت گرادیان (GB و XGBoost) بیشترین دقت (۰/۹۴) و کمترین خطا (خطای MAE به ترتیب: ۰/۰۰۰۴۱۵ و ۰/۰۰۰۴۳۷) را پیش‌بینی مقدار ماسه عبوری در واحد سطح ارائه دادند و می‌توانند عملکرد توری‌های مختلف را با دقت زیادی پیش‌بینی کنند. درنهایت بر اساس مقایسه‌ی نموداری، مدل GB به دلیل مطابقت بهتر مقادیر پیش‌بینی با داده‌های واقعی به‌عنوان مدل برگزیده انتخاب شد.

۲. ویژگی‌های اندازه‌ی عرض شکاف توری و اختلاف فشار بسته‌ی ماسه، تأثیرگذارترین عوامل در تعیین میزان

ماسه عبوری آزمایش محسوب می‌شوند. افزایش اندازه‌ی عرض شکاف توری اثر بسزایی در افزایش میزان ماسه‌ی عبوری دارد. افزایش فشار اعمال‌شده بر نمونه‌ی ماسه نیز باعث کاهش قابل‌توجه میزان ماسه‌ی عبوری می‌شود.

۳. دبی جریان سیال نیز یک عامل مؤثر در عملکرد توری محسوب می‌شود به‌طوری‌که با افزایش دبی آزمایش، ماسه‌ی عبوری بیشتری انتظار می‌رود.

۴. در توزیع اندازه ذرات ماسه، افزایش fine و UC باعث افزایش میزان ماسه عبوری می‌شود، درحالی‌که افزایش SC اثر معکوس دارد. همچنین با افزایش قطر دانه‌ها در نمونه، انتظار می‌رود که میزان ماسه عبوری کاهش می‌یابد.

۵. نوع طراحی لوله مشبک‌کاری شده و اندازه عرض شکاف‌های آن، اهمیت زیادی در بهینه‌سازی میزان ماسه‌ی عبوری دارد و باید با توجه به توزیع اندازه ذرات ماسه انتخاب شود که برای انتخاب مناسب توری می‌توان از مدل ماشینی استفاده کرد.

۶. به‌کارگیری مدل‌های ماشینی در زمینه‌ی پیش‌بینی تولید ماسه در چاه‌های نفت و گاز، می‌تواند در بهینه‌سازی تکمیل چاه‌ها کمک کند.

فهرست علائم اختصاری

معنی فارسی	عبارت کامل انگلیسی	علامت اختصاری
آزمون نگهداشت ماسه	<i>Sand Retention Test</i>	<i>SRT</i>
اعتبارسنجی متقابل	<i>Cross Validation</i>	<i>CV</i>
تعداد شکاف در هر ستون	<i>Slots Per Column</i>	<i>SPC</i>
مساحت ناحیه باز جریان	<i>Open Flow Area</i>	<i>OFA</i>
شبکه عصبی مصنوعی	<i>Artificial Neural Network</i>	<i>ANN</i>
یادگیری ماشین	<i>Machine Learning</i>	<i>ML</i>
تقویت گرادیان	<i>Gradient Boosting</i>	<i>GB</i>
تقویت گرادیان شدید	<i>Extreme Gradient Boosting</i>	<i>XGBoost</i>
جنگل تصادفی	<i>Random Forest</i>	<i>RF</i>
درخت تصمیم	<i>Decision Tree</i>	<i>DT</i>
نزدیک‌ترین همسایه	<i>K-Nearest Neighbors</i>	<i>KNN</i>
ماشین بردار پشتیبان	<i>Support Vector Machine</i>	<i>SVM</i>
پرسترون چندلایه	<i>Multi-Layer Perceptron</i>	<i>MLP</i>
رگرسیون چندجمله‌ای	<i>Polynomial Regression</i>	<i>Poly Reg</i>
توزیع اندازه ذرات	<i>Particle Size Distribution</i>	<i>PSD</i>
ضریب یکنواختی	<i>Uniformity Coefficient</i>	<i>UC</i>
ضریب جور شدگی	<i>Sorting Coefficient</i>	<i>SC</i>
ریشه میانگین مربعات خطا	<i>Root Mean Square Error</i>	<i>RMSE</i>
میانگین خطای مطلق	<i>Mean Absolute Error</i>	<i>MAE</i>
ضریب تعیین	<i>Coefficient of Determination</i>	R^2
واریانس تبیین شده	<i>Explained Variance</i>	<i>EV</i>

- مراجع:
- A, Gupta T, Goel PN, Kumari S. Sand screens for controlling sand production from hydrocarbon wells: A mini-review. *Geoenergy Science and Engineering*. 2024;240:213040.
- [1]. Zare-Reisabadi M, Safari-Beidokhti M, Kaffash A, Bataee M. Sand production prediction and well completion optimization. *International Journal of Petroleum and Geoscience Engineering (IJPGE)*. 2014;2(4):361-74.
- [2]. Kumar S, Kumar G, Krishna S, Kumar
- [3]. Velayati A, Roostaei M, Fattahpour V, Mahmoudi M, Nouri A, Alkoush A, et al., editors. *Design Optimization of Slotted*



- OTC.
- [11]. Reyes R, Sipi A, editors. Optimal sand control design & technique selection: a simplified practical guidance tool. Offshore Technology Conference Asia; 2018: OTC.
- [12]. Ballard T, Beare S, editors. An investigation of sand retention testing with a view to developing better guidelines for screen selection. SPE International Conference and Exhibition on Formation Damage Control; 2012: SPE.
- [13]. Ballard T, Beare SP, editors. Sand retention testing: the more you do, the worse it gets. SPE International Conference and Exhibition on Formation Damage Control; 2006: SPE.
- [14]. Hassani H, Silva ES. Big Data: a big opportunity for the petroleum and petrochemical industry. OPEC Energy Review. 2018;42(1):74-89.
- [15]. Chaudhary S, Yadav S, Kushwaha S, Shahi SRP. A brief review of machine learning and its applications. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology. 2020;12(SUP 1):218-23.
- [16]. Ray S, editor A quick review of machine learning algorithms. 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon); 2019: IEEE.
- [17]. Razak NNA, Abdulkadir SJ, Maoinsar MA, Shaffee SNA, Ragab MG. One-dimensional convolutional neural network with adaptive moment estimation for modelling of the sand retention test. Applied Sciences. 2021;11(9):3802.
- [18]. Ngwashi AR, Ogbe DO, Udebhulu DO, editors. Evaluation of machine-learning tools for predicting sand production. SPE Nigeria Annual International Conference Liner Completions in Cased and Perforated Wells: A Numerical Skin Model. SPE International Heavy Oil Conference and Exhibition; 2018: SPE.
- [4]. Wang C, Pang Y, Mahmoudi M, Haftani M, Salimi M, Fattahpour V, Nouri A. A set of graphical design criteria for slotted liners in steam assisted gravity drainage production wells. Journal of Petroleum Science and Engineering. 2020;185:106608.
- [5]. Guo Y, Nouri A, Nejadi S. Effect of slot width and density on slotted liner performance in SAGD operations. Energies. 2020;13(1):268.
- [6]. Roostaei M, Cespedes EAM, Uzcátegui AA, Soroush M, Hosseini SA, Izadi H, et al. Optimization of slotted liner in rubiales field: trade-off between sand control, flow performance, and plugging tendency. SPE Journal. 2021; 26(03)30-1110.
- [7]. Kaiser T, Wilson S, Venning L. Inflow analysis and optimization of slotted liners. SPE Drilling & Completion. 2002;17(04):200-9.
- [8]. Abduljabbar A, Amadi A, Mohyaldinn ME, Ridha S, Younis O, Alakbari FS. Sand screens application and performance for sand control: A review of selection criteria, screen materials, and causes of failure. Heliyon. 2024;10 (10).
- [9]. Haftani M, Kotb O, Nguyen PH, Wang C, Salimi M, Nouri A. A Novel sand control testing facility to evaluate the impact of radial flow regime on screen performance and its verification. Journal of Petroleum Science and Engineering. 2020;195:107903.
- [10]. Wu B, Choi S, Feng Y, Denke R, Barton T, Wong C, et al., editors. Evaluating sand screen performance using improved sand retention test and numerical modelling. Offshore technology conference Asia; 2016:

- and management in oil and gas wells: a review. Iranian Gas Engineering Journal. 2025;12(1):54-69. [In Persian]
- [26]. Toro C, Pietrangeli G, Lara J, Pineda M, Mesa S, editors. Innovative Remediation System for Mature-Heavy Crude Oil Fields Completed with Slotted Liners and Screens. Offshore Technology Conference Brasil; 2025: OTC.
- [27]. Xie J, editor Slotted liner design optimization for sand control in SAGD wells. SPE Thermal Integrity and Design Symposium; 2015: SPE.
- [28]. Mahmoudi M, Fattahpour V, Wang C, Kotb O, Roostaei M, Nouri A, et al., editors. A Large-Scale Sand Retention Test Facility for Evaluation and Selection of Optimal Standalone Sand Screen for Injection Wells in Thermal Operations. SPE International Conference and Exhibition on Formation Damage Control; 2018: SPE.
- [29]. Mayorga Cespedes EA, Roostaei M, Uzcátegui AA, Soroush M, Izadi H, Hosseini SA, et al., editors. Sand control optimization for rubiales field: trade-off between sand control, flow performance and mechanical integrity. SPE Latin America and Caribbean Petroleum Engineering Conference; 2020: SPE.
- [30]. Ochmann L, Joseph R. Development of a sand retention test set-up focusing on the measurement of produced sand. OIL GAS European Magazine. 2021;2:25-32.
- [31]. Fattahpour V, Roostaei M, Hosseini SA, Soroush M, Berner K, Mahmoudi M, et al. Experiments with stand-alone sand-screen specimens for thermal projects. SPE Drilling & Completion. 2021;36(01):188-207.
- [32]. Mahmoudi M, Fattahpour V, Nouri A, Yao T, Baudet BA, Leitch M, Fermaniuk B, editors. New criteria for slotted liner design and Exhibition; 2021: SPE.
- [19]. Asfha DT, Latiff AHA, Otchere DA, Tackie-Otoo BN, Babikir I, Rafi M, et al. Mechanisms of sand production, prediction—a review and the potential for fiber optic technology and machine learning in monitoring. Journal of Petroleum Exploration and Production Technology. 2024;14(10):2577-616.
- [20]. Azad M, Zargar G, Arabjamaloei R, Hamzei A, Ekramzadeh M. A new approach to sand production onset prediction using artificial neural networks. Petroleum science and technology. 2011;29(19):1975-83.
- [21]. Hamid ZHA, Hernandez N, Baillard F, Looi T, editors. Sand Production and Control Benchmarking through Unstructured Data Analysis with Machine Learning in the North Sea. 84th EAGE Annual Conference & Exhibition; 2023: European Association of Geoscientists & Engineers.
- [22]. Laoufi H, Megherbi Z, Zeraibi N, Merzoug A, Ladmia A, editors. Selection of sand control completion techniques using machine learning. ARMA/DGS/SEG International Geomechanics Symposium; 2022: ARMA.
- [23]. Song J, Li Y, Liu S, Xiong Y, Pang W, He Y, Mu Y. Comparison of machine learning algorithms for sand production prediction: an example for a gas-hydrate-bearing sand case. Energies. 2022;15(18):6509.
- [24]. Laoufi H, Megherbi Z, Zeraibi N, Merzoug A, Ladmia A, editors. Selection of Sand Control Completion Techniques Using Machine Learning. International Geomechanics Symposium; 2022. ARMA-IGS-2022-118.
- [25]. Khodashenas M, Mohammadzadeh Shirazi M, Shahsavani B. Application of machine learning in sand production prediction





- [40]. Uyeh DD, Iyiola O, Mallipeddi R, Asem-Hiablie S, Amaizu M, Ha Y, Park T. Grid search for lowest root mean squared error in predicting optimal sensor location in protected cultivation systems. *Frontiers in Plant Science*. 2022;13:9202.
- for heavy oil thermal production. *SPE Thermal Integrity and Design Symposium*; 2016: SPE.
- [33]. Mahmoudi M, Fattahpour V, Nouri A, Leitch M, editors. An experimental evaluation of pore plugging and permeability reduction near SAGD sand control liners. *SPE Canada Heavy Oil Conference*; 2017: SPE.
- [34]. Devere-Bennett N, editor Using prepack sand-retention tests (SRT's) to narrow down liner/screen sizing in SAGD wells. *SPE Thermal Integrity and Design Symposium*; 2015: SPE.
- [35]. Fischer C, Hamby H, editors. A novel approach to constant flow-rate Sand retention testing. *SPE International Conference and Exhibition on Formation Damage Control*; 2018: SPE.
- [36]. Ismail NI, Kuang S, Zheng E, Yu A. Modeling and analysis of fluid rheology effect on sand screen performance. *Powder Technology*. 2022;411:117961.
- [37]. Cabello-Solorzano K, Ortigosa de Araujo I, Peña M, Correia L, J. Tallón-Ballesteros A, editors. The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. *International conference on soft computing models in industrial and environmental applications*; 2023: Springer.
- [38]. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*. 2021;7:e623.
- [39]. Burzykowski T, Geubbelmans M, Rousseau A-J, Valkenborg D. Validation of machine learning algorithms. *American journal of orthodontics and dentofacial orthopedics*.

Prediction of Sand Passage Rate Through Slotted Liner Using Machine Learning Algorithms

Mohammad Elyas Khodashenas*

M.Sc. Student, Petroleum Engineering, School of Chemical Engineering, Oil and Gas, Shiraz University, Shiraz, Iran

ARTICLE INFO

RESEARCH ARTICLE

Article History:

Received: 16 May 2026

Revised: 20 June 2026

Accepted: 20 June 2026

Keywords:

Sand retention test

Well completion

Sand production

Slotted liner

Machine learning

ABSTRACT

Sand production from oil and gas wells reduces productivity and causes damage to production equipment. To control this phenomenon, mechanical sand screens, particularly slotted liners, are employed, and the performance of these screens is determined through the analysis of sand retention test results. In this study, machine learning algorithms were utilized to predict the amount of sand passing through the unit surface area of slotted liner coupons. To this end, the results of 519 sand retention test samples reported in previous studies were compiled as a dataset, encompassing various screen structures and designs, sand particle size distributions, and experimental conditions. The final dataset consisted of 499 preprocessed data samples. The models were constructed using the following algorithms: GB, XGBoost, RF, DT, KNN, SVM, MLP, and Polynomial Regression. The results demonstrated that the GB algorithm, with the highest accuracy ($R^2 = 0.94$) and the lowest error, alongside the XGBoost algorithm, delivered the best predictive performance. Furthermore, it was found that slot width, pressure differential, and flow rate are the most influential factors affecting the amount of produced sand. Additionally, the screen design type and slot width must be selected in accordance with the particle size distribution of the sand. The machine learning model developed in this research offers advantages such as enhanced prediction accuracy compared to previous methods, providing an interpretable model of screen performance, and increasing the speed of selection, thereby contributing to the control of sand production in wells.

DOR: [20.1001.1.2605.1125.1405.02.26](https://doi.org/10.2605/1125.1405.02.26)

How to cite this article

M.E. Khodashenas, Prediction of Sand Passage Rate Through Slotted Liner Using Machine Learning Algorithms. *Iranian Journal of Gas Engineering*. 2025 13(1): 103-118. (https://ijge.irangi.org/article_738768.html)

* Corresponding Author.

E-mail address: mohammadelyas@hafez.shirazu.ac.ir, (M.E. Khodashenas).

Available online 21 June 2026

25885-5251/© 2014 The Authors. Published by Iranian Gas Institute.

This is an open access article under the CC BY license. (<https://creativecommons.org/licenses/by/4.0>)

