

داده کاوی خصوصیات مکانیکی بتن های حاوی رس های کلسینه شده به عنوان مواد جایگزین در بتن

علی حسین غانمی *

کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه تربیت دبیر شهید رجایی.

امیر طریقت

دانشیار، دانشکده مهندسی عمران، دانشگاه تربیت دبیر شهید رجایی.

چکیده

بتن به دلیل خواص عالی، هزینه کم و در دسترس بودن گسترده، پرمصرف ترین محصول ساختمانی در جهان است. با توجه به افزایش جمعیت جهان، میزان تولید سیمان به عنوان یکی از مواد اصلی مورد استفاده در بتن افزایش یافته است. تولید سیمان باعث انتشار گاز دی اکسید کربن شده که باعث افزایش آلودگی محیط زیست می شود. یکی از روش ها موجود برای جلوگیری از افزایش آلودگی، استفاده از مواد جایگزین به جای سیمان است. از مهمترین مواد جایگزین که در سالیان اخیر مورد استفاده قرار گرفته، می توان به سیمان LC3 اشاره کرد. این نوع سیمان با کاهش مقدار کلینکر سیمان، نیاز به سوخت های فسیلی را کاهش داده که در نتیجه انتشار دی اکسید کربن کاهش می یابد. علاوه بر آن، یکی از خواص مکانیکی مهم بتن، مقاومت فشاری آن بوده که تخمین مقدار آن با توجه به زیاد بودن پارامترهای موجود پیچیده است. در نتیجه از روش های آزمایشگاه که پرهزینه است، استفاده می شود که دارای خطا می باشد. در این مقاله از روش های که بر مبنای درخت تصمیم توسعه می یابند استفاده شده است و عملکرد آن ها با استفاده از الگوریتم ژنتیک بهبود یافته است. دو روش LightGBM و XGBoost با دقت های برابر ۰/۹۵۸ در پیش بینی مقاومت فشاری بتن، عملکرد بهتری نسبت به روش های درخت تصمیم و جنگل تصادفی با دقت های ۰/۹۱ و ۰/۹۳۲ نشان داده اند. همچنین میزان عملکرد پارامترهای ورودی در پیش بینی مقاومت فشاری ارائه شده و یک مجموعه داده جدید مورد تست قرار می گیرد که صحت سنجی روش های ارائه شده بررسی شود.

واژه های کلیدی: بتن، یادگیری ماشین و داده کاوی، سیمان LC3، مواد جایگزین سیمانی، الگوریتم های ارتقای.

* نویسنده مسئول: shayan.ghanemi@sru.ac.ir

۱- مقدمه

منابع بسیار فراوانی است. امروزه با توجه به تقاضای جهانی برای کاهش استفاده از زغال سنگ برای تأمین انرژی به دلیل اثرات مضر آن بر روی محیط زیست، منابع خاکستری بادی در حال کاهش است و این امکان وجود دارد که در آینده‌ای نزدیک این منابع موجود نیز تمام شوند [4].

بر خلاف خاکستر بادی، منابع رس بسیار فراوان است و همچنین دمای مورد نیاز برای عمل کلسینه شده آن نیز توسط سوخت زیستی (bio-fuel) قابل تأمین است. در نتیجه مزایا بیشتری در استفاده از رس به جای سایر مواد سیمانی موجود است. یک مطالعه بر روی رس کلسینه نشان می‌دهد که استفاده از آن به جای ۲۰ تا ۳۰ درصد سیمان پرتلند، میزان انتشار کربن دی اکسید را از ۱۸ تا ۲۷ درصد کاهش می‌دهد. در نتیجه استفاده از رس کلسینه شده، به خصوص سیمان LC3 که باعث کاهش کلینکر مصرفی می‌شود، باعث کاهش انتشار دی اکسید کربن می‌شود [5]. همچنین ترکیب سیمان LC3 به همراه سیمان با ۵۰ درصد کلینکر به صورت موفقیت آمیزی در هند و کوبا مورد استفاده قرار گرفته که نشان‌دهنده خواص مشابهی نظیر سیمان پرتلند بوده است [6,7]. یک آنالیز دقیق توسط موسسه فناوری هند نشان می‌دهد که سیمان LC3 به همراه ۵۰ درصد کلینکر در مقایسه با سیمان پرتلند و سیمان پوزولانی به به ترتیب باعث کاهش دی اکسید کربن به مقدار ۳۰ و ۱۰ درصد می‌شود [4].

مقاومت فشاری یکی از مهم‌ترین خواص مکانیکی بتن بوده که ویژگی‌های آن تابع متغیرهای زیادی است. تخمین مقاومت فشاری بتن یک مشکل پیچیده غیر خطی برای مهندسی ساختمان است که برای به دست آوردن مقدار آن از روش‌های آزمایشگاهی استفاده می‌شود. پیش‌بینی مقاومت فشاری بتن قبل از جایگذاری آن، این اجازه را به مهندسان می‌دهد که برنامه‌ریزی و کنترل کیفیت بتن را بهبود ببخشند. علاوه بر این، یک بتن خوب پیش‌بینی شده می‌تواند باعث صرفه‌جویی در مصرف زمان و افزایش سرعت ساخت و ساز (به طور عمده در ساخت و سازهای بزرگ با مراحل بتن‌ریزی زیاد) شود [8-10].

امروزه با توجه به فراگیر شدن روش‌های هوش مصنوعی می‌توان از روش‌های یادگیری ماشین و داده‌کاوی که قابلیت یادگیری از

بتن به دلیل سهولت در استفاده، خواص عالی، هزینه نسبتاً کم و در دسترس بودن گسترده، پر مصرف‌ترین محصول ساختمانی منفرد در جهان است. بتن عمدتاً از سه جز اصلی سیمان، آب و سنگدانه تشکیل شده است که تولید سیمان از بین آنها نه تنها بیشترین هزینه را دارد، بلکه فرایند تولید آن باعث انتشار دی اکسید کربن می‌شود. از سوی دیگر افزایش جمعیت جهان و شهرنشینی باعث افزایش تولید بتن شده که در نتیجه آن میزان تولید سیمان و انتشار دی اکسید کربن افزایش یافته است.

طبق توافق انجام شده در پاریس که توسط سازمان جهانی انرژی ارائه شده است، میزان افزایش گرمای جهانی در این قرن باید محدود به ۲ درجه سانتیگراد باشد. برای دستیابی به این توافق، میزان دی اکسید کربن منتشر شده از تولید سیمان باید ۲۴ درصد کاهش پیدا کند که همانطور که گفته شده این میزان در حال افزایش است. طبق اعلام سازمان انرژی جهانی و تحقیقات انجام شده، یکی از راه‌حل‌های موجود استفاده از مواد جایگزین به جای سیمان بوده که باعث کاهش میزان کلینکر مورد نیاز برای سیمان می‌شود. تولید کلینکر برای استفاده در سیمان، نیازمند انرژی حرارتی زیادی است که از طریق سوخت‌های فسیلی تولید شده و در نتیجه آن دی اکسید کربن تولید می‌شود [1].

مواد جایگزین زیادی نظیر خاکستر گیاهان، خاکستر بادی، پوزولان‌های طبیعی و سنگ آهک می‌توانند به عنوان جایگزین سیمان استفاده شوند اما این مواد واکنش پذیری ناچیزی دارند و قابلیت کاهش کلینکر را ندارند. در طول سالیان گذشته سیمان‌های ساخته شده از رس کلسینه شده مورد مطالعه قرار گرفته‌اند و اخیراً مطالعات گسترده درباره ساختار آنها یک پتانسیل قابل توجه در ترکیب سنگ آهک و رس کلسینه شده به عنوان جایگزین کلینکر نشان می‌دهد.

بر اساس سازمان انرژی جهانی فاکتور کلینکر فعلی ۶۵ درصد بوده که سیمان‌های متشکل از رس کلسینه شده و سنگ آهک بدون تأثیر گذاشتن بر خصوصیات سیمان، میزان کلینکر آن را به ۵۰ درصد کاهش می‌دهد. این نوع سیمان‌ها، سیمان LC3^۱ نامیده می‌شود [۲،۳]. همچنین در مقایسه با خاکستر بادی، رس دارای

^۱ Limestone Calcined Clay Cement

خطی، رگرسیون چند درجه، درخت تصمیم و جنگل تصادفی و... در زبان برنامه نویسی R استفاده کرده و مقاومت فشاری بتن را پیش بینی کرده اند.

به طور کلی در سالیان گذشته، روش های مختلفی از یادگیری ماشین برای پیش بینی مقاومت فشاری بتن حاوی مواد جایگزین سیمانی استفاده شده است. در جدول ۱ خلاصه ای از کاربرد روش های یادگیری ماشین و نتایج آنها ارائه شده است. در این مقاله از روش های یادگیری ماشین XGBoost و LightGBM که الگوریتم های جدیدتری نسبت الگوریتم های موجود در جدول ۱ است، استفاده می شود و عملکرد آنها با روش های درخت تصمیم و جنگل تصادفی مقایسه می شود. همچنین بیشتر مطالعات گذشته بر روی مواد جایگزین سیمانی نظیر خاکستر بادی و سرباره آهن گدازی بوده است. در این مقاله علاوه بر مواد ذکر شده، داده های شامل رس کلسینه، کائولینه و میکروسیلیس جمع آوری شده است که یک مدل جامع بدست آید.

روی داده های موجود را دارند، استفاده کرد. دو روش یادگیری ماشین و داده کاوی روی یکدیگر همپوشانی دارند با این تفاوت که یادگیری ماشین بر پیش بینی خواص شناخته شده و داده کاوی بر کشف خواص ناشناخته متمرکز است [11].

در سال ۲۰۱۰ دیا و همکاران [12] مقاومت فشاری بتن های توانمند را با استفاده از مدل های بر پایه درخت تصمیم پیش بینی کرده اند. در سال ۲۰۱۸ کیم و همکاران [13]، با استفاده از روش های مختلف در رگرسیون به پیش بینی مقاومت فشاری بتن حاوی خاکستر بادی، کوره آهن گدازی، سیمان، ریزدانه، سنگدانه و روان کننده پرداخته اند و روش های استفاده شده را به هم مقایسه کرده اند. در سال ۲۰۱۱ چو و همکاران [14] از روش هایی نظیر شبکه عصبی (ANN)، ماشین بردار پشتیبان (SVM) و روش درختی (Ensemble Bagging) برای پیش بینی مقاومت فشاری بتن های حاوی Blast Furnace Slag استفاده کردند. در سال ۲۰۱۹ دیروسیو و همکاران [15] از مدل های رگرسیون

جدول ۱- خلاصه تحقیقات پیشین

تحقیقات پیشین	تعداد داده ها	الگوریتم استفاده شده	خروجی R^2	مواد جایگزین سیمانی
یه [35]	۷۲۷	شبکه عصبی	۰/۹۱۴	خاکستر بادی و سرباره آهن
		رگرسیون خطی	۰/۵۷۴	گدازی
یه و لین [39]	۱۱۹۶	شبکه عصبی	۰/۹۳۳	خاکستر بادی و سرباره آهن
		الگوریتم ژنتیک و درخت تصمیم	۰/۸۶۷	گدازی
سارادمیر [10]	۱۹۵	شبکه عصبی	-	رس کائولینه و میکروسیلیس
دیا و همکاران [12]	۳۰۰	شبکه عصبی	۰/۶۲۵	خاکستر بادی و سرباره آهن
		رگرسیون خطی	۰/۴۹۱	گدازی
		درخت MP5	۰/۷۸۷	
چو و همکاران [14]	۱۰۳۰	شبکه عصبی	۰/۹۰۹	خاکستر بادی و سرباره آهن
		ماشین بردار پشتیبان	۰/۸۸۵	گدازی
		درخت رگرسیون Bagging	۰/۸۹	
اردال و همکاران [40]	۱۰۳۰	شبکه عصبی	۰/۹۰۸	خاکستر بادی و سرباره آهن
		ترکیب شبکه عصبی و Bagging	۰/۹۲۷	گدازی
چنگ و همکاران [41]	۱۰۳۰	آدا بوست (AdaBoost)	۰/۹۸	خاکستر بادی و سرباره آهن

یکی از چالش‌های موجود در استفاده از روش‌های یادگیری ماشین، تنظیم هایپرپارامترهای (پیدا کردن بهترین ترکیب هایپرپارامترها) آن است. الگوریتم‌های یادگیری ماشین حاوی چندین هایپرپارامتر و پارامتر هستند که تفاوت اصلی آن‌ها در این است که هایپرپارامترها در طول فرآیند آموزش مدل تاثیرگذار هستند و مدل توانایی به دست آوردن مقدار آن‌ها را ندارد. در نتیجه از روش‌های دستی یا از روش‌های بهینه کردن هایپرپارامترها برای بدست آوردن بهترین ترکیب از این مقادیر استفاده می‌شود [16]. در این مقاله برای بهبود عملکرد الگوریتم‌های استفاده شده از روش الگوریتم ژنتیک استفاده شده که یکی از معروف‌ترین روش‌های بهینه‌یابی است.

۲- پیش پردازش داده

در فرآیند پیش پردازش داده، جمع‌آوری، پاکسازی داده و آنالیز داده برای بدست آوردن نتایج بهتر مورد استفاده قرار گرفته است.

جدول ۲- توزیع آماری پارامترهای مجموعه داده

انحراف معیار	میانگین	ماکزیمم	مینیمم	
۲۶/۹۷	۸/۰۷	۱۶۵/۲	۰	رس کلسینه
۱۰۵/۲۲	۳۲۶/۸۳	۵۸۳	۱۱۶	سیمان
۱۰/۷	۲/۱۵	۷۵	۰	میکرو سیلیس
۲۶/۶۵	۱۱/۴۸	۱۷۴/۹	۰	رس کائولینه
۵۶/۴۹	۳۳/۱۵	۲۰۰/۱	۰	خاکستر بادی
۷۴/۹۲	۴۱/۹۱	۳۹۴/۴	۰	سرباره آهن‌گذاری
۱۰۲/۶۶	۷۳۹/۶۳	۹۹۲/۶	۳۹۰	ریزدانه
۹۹/۴۱	۱۰۲۱/۶۸	۱۳۰۰	۸۰۱	درشت دانه
۰/۱۱	۰/۴۴	۰/۷	۰/۲۴	نسبت آب به مواد چسبنده
۵۶/۲۲	۴۴/۳	۳۶۵	۱	سن

پاکسازی داده صرفاً به معنای حذف کردن اطلاعات برای ایجای فضای اضافی برای داده‌های جدید نیست بلکه به معنای یافتن راهی برای به حداکثر رساندن دقت در مجموعه داده بدون حذف اطلاعات است. پاکسازی داده شامل اقدامات بیشتری نسبت به حذف داده، مانند رفع خطاهای دستوری و املائی، استانداردسازی مجموعه داده و تصحیح اشتباهاتی نظیر زمینه‌های خالی، کدهای مفقود شده و شناسایی داده‌های تکراری است.

پاکسازی داده یکی از عناصر اساسی میانی علوم داده به شمار می‌رود زیرا در فرآیند تحلیل و کشف پاسخ‌های معتبر نقش مهمی ایفا می‌کند. با توجه تحقیقات انجام شده، آماده‌سازی داده ۸۰ درصد از زمان انجام یک پروژه را در برمی‌گیرد که نشان دهنده اهمیت آن است [42]. در ادامه فرآیند پاکسازی داده به شرح زیر انجام شده که باعث کاهش تعداد داده از ۱۹۴۰ به ۱۶۸۷ شده است:

-چک کردن واحدها: اولین کار در انجام پاکسازی داده‌ها،

- بدست آوردن اطلاعاتی نظیر مقدار واریان، مقدار ماکزیمیم، مقدار مینیمم و میزان توزیع داده ها.
- بدست آوردن رابطه خطی بین پارامترهای ورودی و پارامتر خروجی بوسیله ماتریس همبستگی.

در جدول ۳ اطلاعات مربوط به ماتریس همبستگی ارائه شده است. ماتریس همبستگی دید اولیه مناسبی در رابطه با پارامترهای ورودی در اختیار ما قرار می دهد. این روش ارتباط خطی مستقیم هر پارامتر ورودی با پارامتر خروجی را به صورت مجزا مورد بررسی قرار می دهد و روش کاملی نیست. اعداد منفی نشان دهنده رابطه عکس با پارامتر خروجی است. همچنین مقادیر بین ۰ تا ۱ است که هرچه قدر به ۱ نزدیکتر باشد نشان دهنده تاثیر بیشتر و هرچه قدر به ۰ نزدیکتر باشد نشان دهنده تاثیر ناچیز روی مقاومت فشاری به عنوان پارامتری خروجی است.

جدول ۳- ماتریس همبستگی

رس کلسینه	-۰/۰۸	سرباره آهن گذاری	-۰/۱۴
رس کائولینه	۰/۳	ریزدانه	-۰/۲۳
سیمان	۰/۵۵	درشت دانه	۰/۱۸
میکروسیلیس	۰/۲۸	سن بتن	۰/۳۱
خاکستر بادی	-۰/۲۶	آب به مواد چسبنده	-۰/۶۷

از آنجا که روش ماتریس همبستگی روش کاملی نیست به بررسی اثر رس کلسینه و رس کائولینه روی مقاومت فشاری با ترسیم شکل آن ها می پردازیم. در شکل ۱ نمودار رس کلسینه و مقاومت فشاری ارائه شده است.

همانطور که مشخص است با توجه به خط ارائه شده در شکل ارائه شده، افزودن رس کلسینه شده باعث افزایش مقاومت فشاری می شود. برای بررسی بهتر، تاثیر سن بتن نیز مورد توجه قرار می گیرد. شکل ۲ میزان مقاومت فشاری را در سنین مختلف بتن در ۴ مقدار مختلف رس کلسینه ارائه داده است. با توجه به شکل، در سنین کم افزایش مقدار رس کلسینه باعث کاهش مقاومت فشاری بتن می شود اما نکته قابل توجه این است که در سنین بالا، این افزایش مقدار، می تواند باعث افزایش مقاومت فشاری بتن شود. همچنین مقدار رس کلسینه اگر یک مقدار بهینه باشد که در اینجا با توجه به شکل ۱۱۰/۱

چک کردن واحد داده ها است. ممکن مقدار بعضی داده ها بر اساس درصد بیان شده باشد که باید همه ی آنها بر اساس واحد اصلی و یکسان همه ی داده ها که کیلوگرم بر مترمکعب است محاسبه و وارد شوند.

- **حذف کردن مشاهدات و ستون ها:** در این مرحله مشاهدات و ستون هایی که اطلاعات کم و ناقصی دارند حذف می شوند. همچنین از بین ستون هایی که تاثیر یکسانی دارند، یک کدام از آنها حذف می شود.

- **داده های محاسبه نشده:** در مجموعه داده جمع آوری شده با یک بررسی ساده می توان متوجه شده که تعدادی از داده ها دارای مقدار NaN هستند که معنی آن Not a Number است. ممکن در برخی مقالات مقدار یک سری از پارامترها محاسبه نشده باشد و بر ما واجب است که اینکار را انجام دهیم. به عنوان مثال پارامتر آب در بعضی مقالات محاسبه نشده است ولی پارامترهای سیمان و آب به سیمان محاسبه شده است.

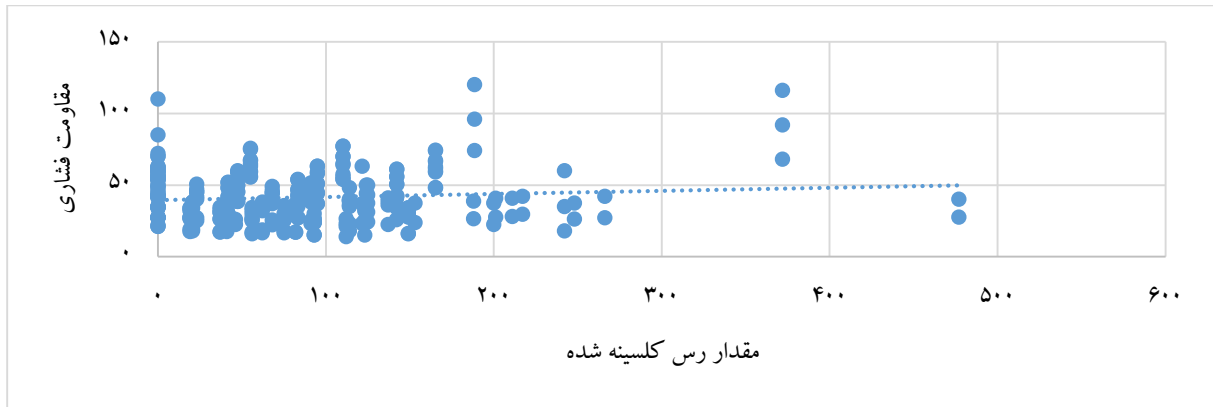
- **داده های گمشده^۱:** در مجموعه داده جمع آوری شده یک سری پارامترها مقداری ندارند یا به عبارت دیگر مقدار آن های گمشده است. تعداد مشاهداتی که همچین شرایطی دارند کم است و حذف کردن این داده ها روش مناسبی نیست و باعث از بین رفتن اطلاعات مفید می شود. در این مقاله از روش میانگین برای جایگزینی این مقادیر استفاده شده است.

- **داده های پرت:** در مجموعه داده، توزیع آماری داده دید مناسبی به ما می دهد. اگر در یکی از پارامترها، تعدادی کمی از داده ها دارای مقادیر بسیار بزرگی یا بسیار کوچک می تواند این مقادیر را حذف کرد یا به عنوان یک گروه جدا در نظر گرفت و بررسی کرد. در این مقاله با توجه به ناچیز بودن تعداد داده های پرت، این مقادیر حذف شده اند.

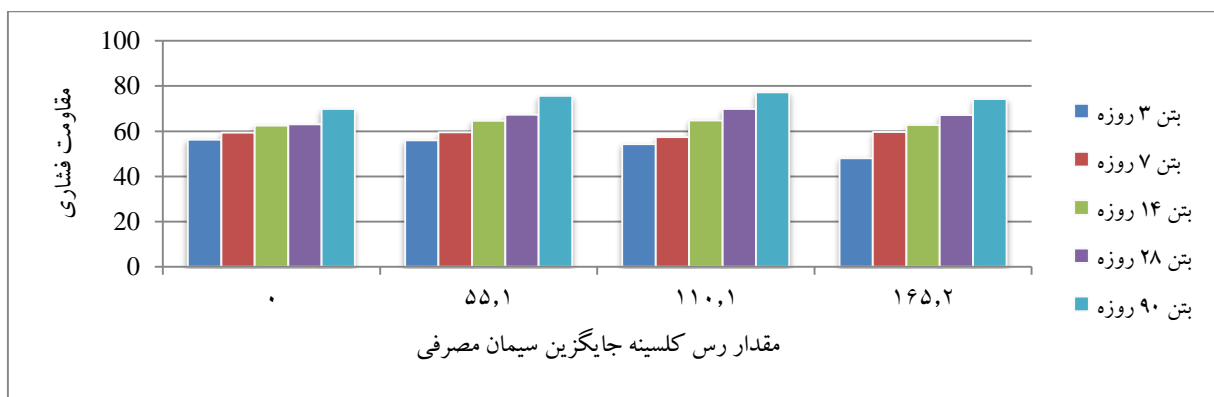
پس از اتمام پاکسازی داده تحلیل و آنالیز داده انجام شده است. هدف از انجام این مرحله به دست آوردن دید بهتر نسبت به متغیرهای موجود در مجموعه داده جمع آوری شده است. به طور کلی این مرحله یک دید کلی در رابطه با ارتباط پارامترهای ورودی نسبت به پارامتر خروجی و توزیع آماری به ما می دهد [42]. بعد از انجام این مرحله اطلاعات زیر بدست می آید:

¹ Missing Value

است، می تواند باعث مقاومت بیشتر فشاری بتن شود. در شکل ۳ میزان مقاومت فشاری بتن در سنین مختلف و در ۸ مقدار مختلف رس کائولینه ارائه شده است. همانطور که قابل مشاهده است، با افزایش مقدار رس کائولینه مقاومت فشاری بتن افزایش می یابد اما بهینه باشد، می تواند باعث افزایش مقاومت فشاری بتن شود.



شکل ۱- نمودار تاثیر رس کلسینه بر روی مقاومت فشاری



شکل ۲- تاثیر جایگزینی رس کلسینه بر روی مقاومت فشاری بتن در سن های مختلف



شکل ۳- تاثیر جایگزینی رس کائولینه بر روی مقاومت فشاری بتن در سن های مختلف

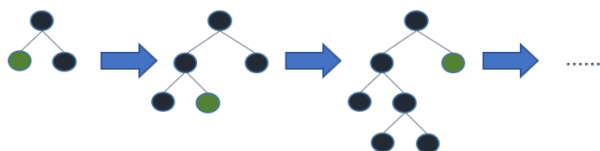
به طور کلی مواد سیلیسی نظیر رس ها با جذب آهک موجود در بتن، ژل ثانویه و سیمان مانند تشکیل می دهد. این ژل فضای خالی بین سنگدانه ها را پر کرده و باعث کاهش تخلخل و افزایش مقاومت فشاری بتن می شود. با این حال اگر میزان مصرف این مواد زیاد باشد با توجه به اتمام آهک موجود در بتن، نقش ماده پرکننده پیدا می کند و باعث کاهش مقاومت فشاری بتن می شود [43].

داده با تعداد داده معمولی عملکرد بسیار مناسبی دارند. در این روش، سرعت آموزش مدل بسیار سریع است و همچنین تنظیم های پارامترهای مدل زمان زیادی نمی گیرد [44].

استراتژی روش boosting به این صورت است که چندین یادگیری ضعیف (weak learner) را با استفاده از تنظیمات مجزا یا تعداد داده های مختلف مورد آموزش قرار می دهد. سپس در انتها تمام یادگیری های ضعیف کنار یکدیگر قرار گرفته و با هم ترکیب می شوند. این عمل باعث افزایش دقت خروجی مدل نهایی می شود [45]. دو تا از بهترین و موثرترین مدل های یادگیری ماشین که بر اساس درخت تصمیم و استراتژی boosting عمل می کنند و در سالیان اخیر کاربرد آنها در صنعت و آکادمیک افزایش یافته، XGBoost و LightGBM است [45]. به طور کلی دو الگوریتم XGBoost و LightGBM مشابه یکدیگر هستند با این تفاوت که استراتژی توسعه دیاگرام درخت تصمیم در روش XGBoost به صورت شکل ۴ و در روش LightGBM به صورت شکل ۵ است.



شکل ۴- نحوه تصمیم گیری در استراتژی رشد هم سطح برگ [44]



شکل ۵- نحوه تصمیم گیری در استراتژی رشد برگ [44]

۳-۴- الگوریتم ژنتیک

الگوریتم ژنتیک یک الگوریتم فرا ابتکاری و زیرمجموعه الگوریتم های تکاملی است. این الگوریتم از نظریه تکاملی داروین استخراج شده است و در مسائل بهینه یابی ریاضی مورد استفاده قرار می گیرد [46].

الگوریتم ژنتیک به طور کلی از چندین کروموزوم تشکیل شده و هر کروموزوم از چندین ژن تشکیل شده است. مجموع کروموزوم های موجود، جمعیت کل را تشکیل می دهد. برای حل

پس از اتمام پیش پردازش داده، داده های موجود در مجموعه داده به دو دسته آموزش و تست تقسیم می شوند. ۲۰ درصد از داده های موجود به داده تست و ۸۰ درصد به داده آموزش اختصاص می یابد. با توجه به جدول ۴ تعداد داده های آموزش و تست مشخص شده است.

جدول ۴- تقسیم بندی داده ها

	آموزش	تست
تعداد داده	۱۳۵۰	۳۳۷

۳- یادگیری ماشین

در این قسمت، روش های یادگیری ماشین استفاده شده در این مقاله توضیح داده می شود.

۳-۱- درخت تصمیم

درخت تصمیم گیری یک از روش های پر کاربرد یادگیری ماشین است و نحوه کار آن به این صورت است که داده ها را به چند زیر مجموعه کوچک تقسیم کرده و همان زمان یک درخت تصمیم توسعه می دهد. بعد از تصمیم گیری نهایی در مورد همه ی زیر مجموعه ها در نهایت دیاگرام درخت تصمیم گیری شکل می گیرد.

۳-۲- جنگل تصادفی

جنگل تصادفی همانند درخت تصمیم گیری یکی از روش های پر کاربرد در حوزه یادگیری ماشین است. همانطور که گفته شد درخت تصمیم گیری براساس کل داده تصمیم گیری می کند ولی در این روش الگوریتم به صورت تصادفی چندین پارامتر و مشاهده را انتخاب کرده و چندین درخت تصمیم گیری تشکیل می شود و در نهایت میانگین نتایج مورد استفاده قرار می گیرد. یکی از مزیت های این روش نسبت به درخت تصمیم این است از بیش برآزش جلوگیری می کند [3].

۳-۳- الگوریتم های Boosting

با توجه به اینکه استفاده از الگوریتم شبکه عصبی در سالیان گذشته افزایش یافته است، اما روش های boosting که زیر مجموعه روش های ensemble learning است همچنان روی مجموعه

مدل بد است. همچنین مقادیر P_i ، T_i و \bar{T} به ترتیب مقدار پیش‌بینی شده توسط مدل، مقدار واقعی و مقدار میانگین پارامتر خروجی مدل (مقاومت فشاری) است.

در جدول ۲ نتایج مدل‌های استفاده شده ارائه شده است که با توجه به آن مدل‌های XGBoost و LightGBM عملکرد بهتری نسبت به درخت تصمیم و جنگل تصادفی دارند. همچنین جنگل تصادفی عملکرد بهتری نسبت به درخت تصمیم دارد. در الگوریتم درخت تصمیم یک فاصله بین نمره آموزش و تست وجود دارد که می‌تواند نشان‌دهنده بیش‌برازش (overfit) باشد که مشکل اصلی این الگوریتم است. جنگل تصادفی که از چندین درخت تصمیم تشکیل شده است توانسته تا حدودی این مشکل را حل کند که همانطور که مشخص است فاصله بین نمره داده‌های آموزش و تست تا حدی کاهش یافته است. دو الگوریتم XGBoost و LightGBM که تقریباً مشابه همدیگر هستند عملکرد مناسبی در داده‌های آموزش و تست داشته‌اند. میزان خطای (MAE و RMSE) روش XGBoost در داده‌های آموزش به ترتیب برابر ۲/۱۲ و ۳/۳۵ که تقریباً برابر با روش LightGBM است. دقت روش XGBoost برابر با ۰/۹۷۳ و در روش LightGBM ۰/۹۷۴ است که نشان می‌دهد این دو روش دقت بالایی دارند. در داده‌های تست که قدرت پیش‌بینی هر الگوریتم را نشان می‌دهد همچنان دو روش LightGBM و XGBoost عملکرد بسیار مناسبی دارند و مشخص است که این دو روش قدرت بالایی در پیش‌بینی مقاومت فشاری بتن دارند. همچنین با توجه به پارامتر SI، عملکرد دو روش XGBoost و LightGBM به دلیل کمتر بودن مقدار پارامتر SI از ۰/۱ عملکرد کلی مدل عالی است.

همچنین برای بررسی صحت‌سنجی مدل‌های آموزش داده شده، یک مجموعه داده از مقاله [5] که نوع ماده جایگزین سیمانی این مقاله رس کلسینه شده است، جمع‌آوری شده و عملکرد مدل‌های استفاده شده در پیش‌بینی مقاومت فشاری بتن مورد ارزیابی قرار می‌گیرد. قابل ذکر است که این مجموعه داده در آموزش مدل‌ها استفاده نشده است ولی توزیع مشابهی از آن‌ها در مجموعه داده اصلی موجود است. جدول ۳ نتیجه عملکرد الگوریتم‌ها را روی این مجموعه داده جدید نشان می‌دهد و با توجه به آن دو مدل XGBoost و LightGBM عملکرد بسیار مناسبی در پیش‌بینی مقاومت فشاری این مجموعه جدید نشان داده‌اند. درخت تصمیم

مشکل پیدا کردن بهترین ترکیب هایپرپارامتر، هر کروموزم نشان دهنده یک هایپرپارامتر است و در ژن‌ها ثبت شده است. برای توسعه ترکیب این کروموزوم‌ها، از روش mutation و crossover استفاده می‌شود. در روش mutation به صورت تصادفی جای ژن‌ها عوض شده و کروموزم جدید تولید می‌شود. همچنین در روش crossover ژن‌های کروموزوم‌های مختلف با یکدیگر ترکیب می‌شوند و کروموزوم جدید تشکیل می‌دهند. در طول این فرآیند یک تابع برازش (fitness function) میزان عملکرد کروموزوم‌ها را نشان می‌دهد و در این فرآیند کروموزوم‌ها با عملکرد بد حذف شده، و عمل توسعه روی کروموزوم‌های با عملکرد خوب انجام می‌شود و به عبارت کلی دانش به نسل جدید انتقال داده می‌شود. این فرآیند اینقدر تکرار می‌شود که به نقطه بهینه کینه دست پیدا کنیم (بهترین ترکیب هایپرپارامترها). در الگوریتم ژنتیک تعداد جمعیت کل ۳۰، تعداد نسل ۵، تعداد اعتبارسنجی مدل نیز برابر ۵ بوده است که مورد استفاده قرار گرفته است.

۴- نتایج الگوریتم‌های یادگیری ماشین

در این قسمت نتایج الگوریتم‌های استفاده شده که با استفاده از الگوریتم ژنتیک بهترین ترکیب هایپرپارامترهای هر کدام بدست آمده ارائه می‌شود. همچنین برای مقایسه عملکرد الگوریتم‌های استفاده شده، از فرمول‌های زیر استفاده می‌شود:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - T_i| \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - T_i)^2}{\sum_{i=1}^n (T_i - \bar{T})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - T_i)^2}{n}} \quad (3)$$

$$SI = RMSE / \bar{T} \quad (4)$$

ه‌طور کلی MAE و RMSE نشان دهنده میزان خطای مدل هستند و هرچه مقدار آن‌ها به صفر نزدیکتر باشد بهتر است. R^2 نشان دهنده دقت مدل است و به عنوان ضریب تعیین شناخته می‌شود و هرچه مقدار آن به ۱ نزدیکتر باشد بهتر است. SI یک پارامتر آماری است که میزان مؤثر بودن مدل را نشان می‌دهد به این صورت که اگر مقدار آن ۰ و ۰/۱ باشد عملکرد مدل عالی، اگر بین ۰/۱ و ۰/۲ باشد عملکرد آن خوب و بیشتر از ۰/۲ باشد عملکرد

عملکرد متوسط و جنگل تصادفی نیز عملکرد خوبی از خود نشان داده‌اند.

جدول ۲- نتایج الگوریتم‌های یادگیری ماشین

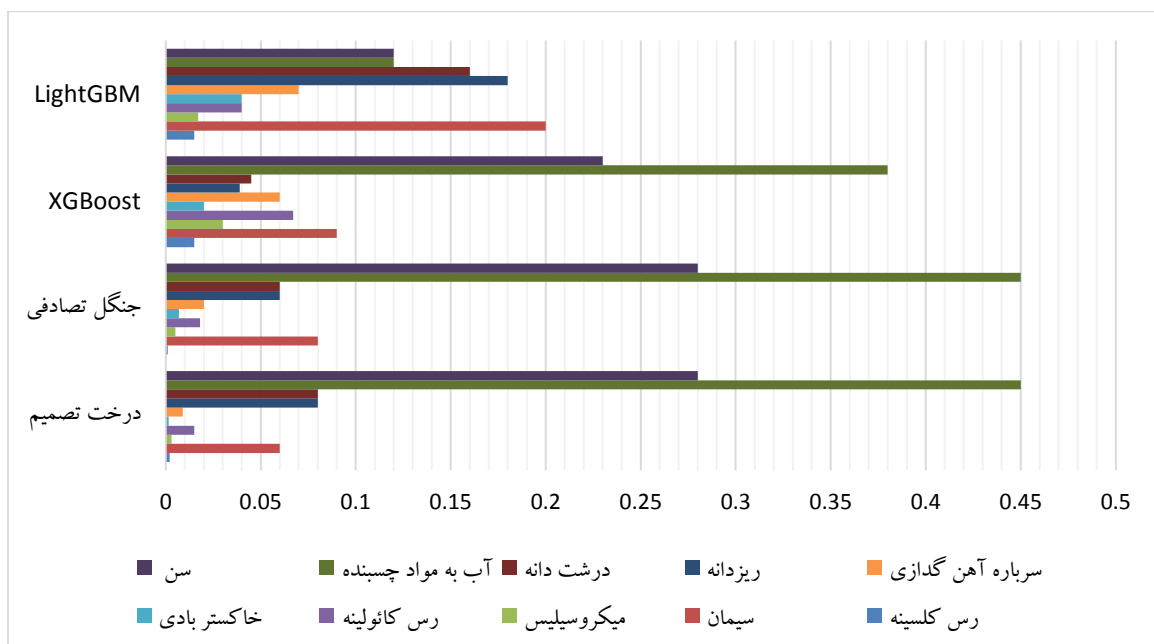
نمره داده‌های تست				نمره داده‌های آموزش				
SI	RMSE	MAE	R^2	SI	RMSE	MAE	R^2	
۰/۱۴	۶/۴۵	۴/۴۲	۰/۹۱	۰/۰۹	۴/۱۶	۳/۰۷	۰/۹۶	درخت تصمیم
۰/۱۱	۵/۲	۳/۷۲	۰/۹۳۲	۰/۰۹	۳/۸۸	۲/۶۱	۰/۹۶۳	جنگل تصادفی
۰/۰۹۷	۴/۲۸	۳/۰۲	۰/۹۵۸	۰/۰۷۸	۳/۳۹	۲/۱۷	۰/۹۷۳	XGBoost
۰/۰۹۷	۴/۲۶	۳	۰/۹۵۸	۰/۰۷۷	۳/۳۵	۲/۱۲	۰/۹۷۴	LightGBM

شده از الگوریتم ژنتیک برای پیدا کردن بهترین ترکیب هایپرپارامترها استفاده می‌شود که نتایج آن در جدول ۴ ارائه شده است. با توجه به جدول ۷، max depth که همان تعداد زیرشاخه‌های موجود در دیاگرام درخت تصمیم است، پیچیدگی دو روش جنگل تصادفی و درخت تصمیم بیشتر از دو روش دیگر است. بقیه هایپرپارامترهای استفاده شده می‌توانند به مدل کمک کنند که دچار بیش برازش نشود. هایپرپارامتر number of estimator که نشان دهنده تعداد درخت‌های بررسی شده است و با توجه به اینکه درخت تصمیم تنها یک درخت کلی روی مدل آموزش می‌دهد این هایپرپارامتر را دارا نمی‌باشد. در آخر با توجه به نتایج جدول، ترکیب این هایپرپارامترها در الگوریتم ژنتیک نتایج بهتری (بهینه) در این الگوریتم‌ها می‌دهد.

جدول ۳- نتایج صحت سنجی

RMSE	MAE	R^2	
۴/۷۴	۳/۷۶	۰/۸۶	درخت تصمیم
۳/۶۳	۲/۶	۰/۹۱	جنگل تصادفی
۲/۸۱	۲/۴۵	۰/۹۵۶	XGBoost
۲/۹۷	۲/۲۸	۰/۹۴۷	LightGBM

روش‌های درختی استفاده شده، میزان تأثیر هر کدام از پارامترهای ورودی بر پیش‌بینی پارامتر خروجی را به ما می‌دهد. شکل ۱ میزان تأثیر هر پارامتر ورودی را در هر مدل نشان می‌دهد. با توجه به شکل دو پارامتر آب به مواد چسبنده و سن بتن بیشترین تأثیر را در پیش‌بینی مقاومت فشاری دارند. همانطور که گفته شد، برای بهبود عملکرد الگوریتم‌های استفاده



شکل ۱- میزان تأثیر گذاری پارامترهای ورودی در هر الگوریتم

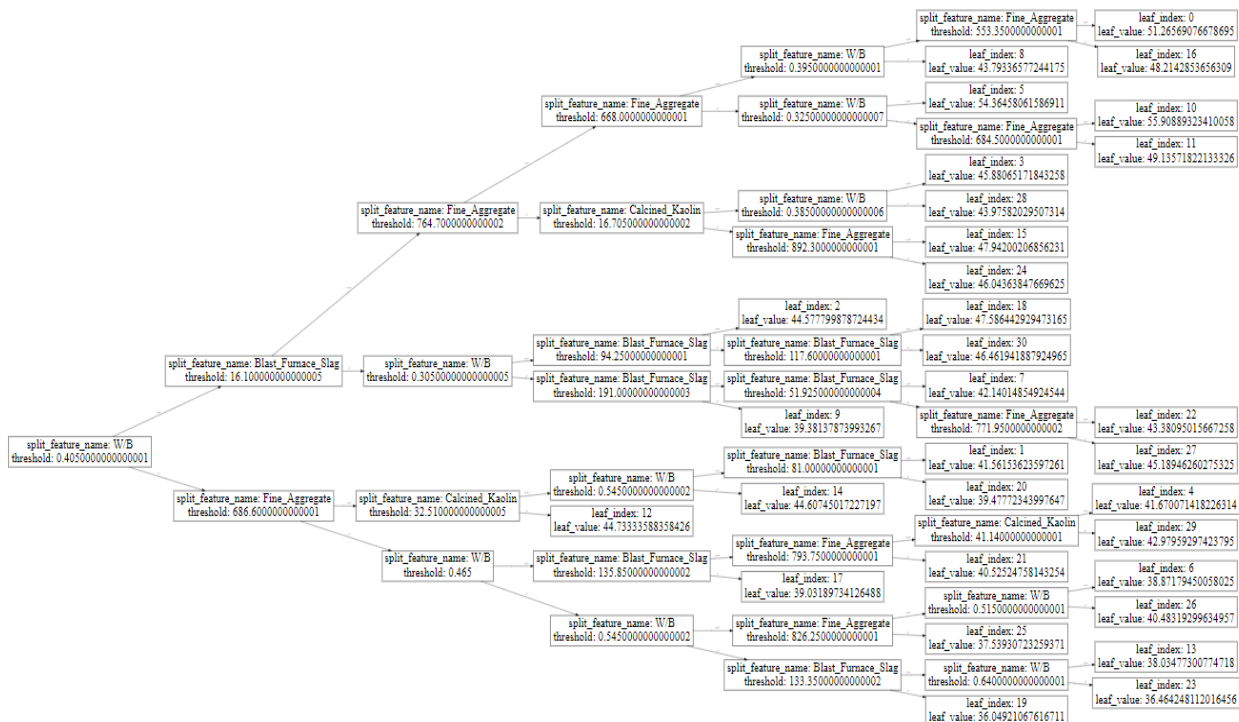
جدول ۴- نتایج بهترین ترکیب های پارامترها به کمک الگوریتم ژنتیک

LightGBM	XGBoost	جنگل تصادفی	درخت تصمیم	
۶	۶	۹	۹	Max depth
-	-	۳	۵	Min sample split
-	-	۲	۱	Min sample leaf
۱۲۶	۱۱۳	۸۳	-	Number of estimator
۰/۳	۰/۱	-	-	Learning rate
۱	۱	-	-	subsample
۰/۵	۰/۵	-	-	Colsample by tree
۴۰	۴۰	-	-	Min child sample
-	۰/۱	-	-	gamma

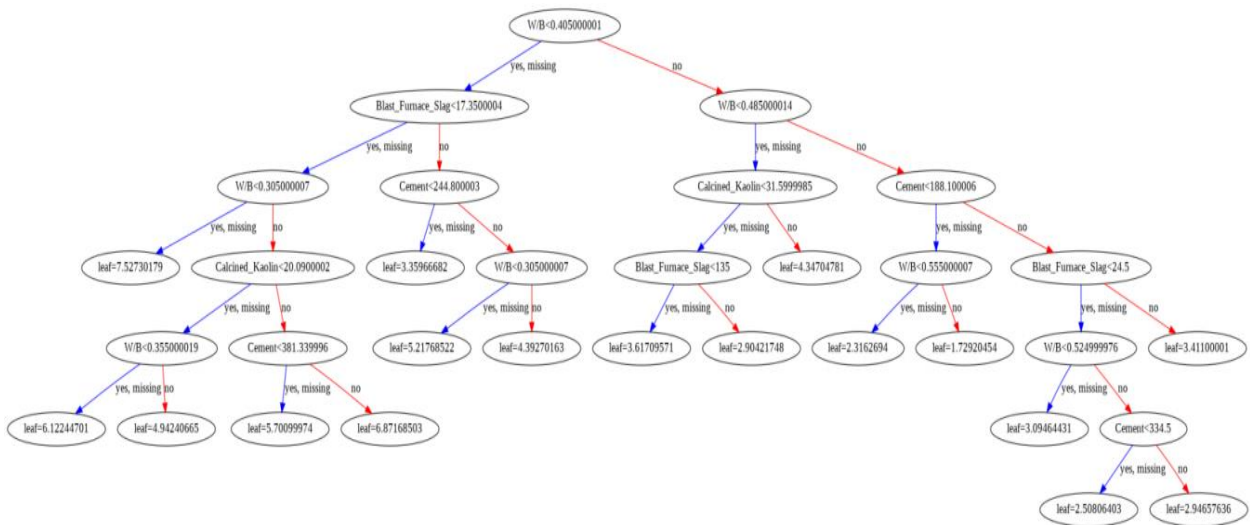
۵- نتیجه گیری

- دو الگوریتم XGBoost و LightGBM عملکرد بهتری نسبت به الگوریتم های درخت تصمیم و جنگل تصادفی دارند. دقت الگوریتم XGBoost و LightGBM در پیش بینی مقاومت فشاری بتن برابر با ۰/۹۵ است. همچنین خطای این دو الگوریتم تقریباً برابر با ۳ و ۴/۲۶ است که بسیار مناسب است. همچنین این دو مدل دارای SI کمتر از ۰/۱ می باشند که نشان می دهد که عملکرد بسیار مناسبی دارند.

در این مقاله از روش های بر پایه درخت تصمیم برای داده کاوی خصوصیات مکانیکی بتن خصوصاً مقاومت فشاری، استفاده شده است. به طور کلی با توجه به الگوریتم های استفاده شده و تحلیل مجموعه داده می توان به نتایج زیر اشاره کرد و همچنین نحوه رشد درخت تصمیم در دو روش XGBoost و LightGBM در شکل ۲ و شکل ۳ که بهترین نتیجه را داشته اند ارائه شده است:



شکل ۲- دیاگرام رشد درخت تصمیم روش LightGBM



شکل ۳- دیاگرام رشد درخت تصمیم روش XGBoost

<https://doi.org/10.1016/j.cemconres.2017.08.017>.
 [5] D. Zhou, R. Wang, M. Tyrer, H. Wong, C. Cheeseman, Sustainable infrastructure development through use of calcined excavated waste clay as a supplementary cementitious material, *J. Clean. Prod.* 168 (2017) 1180–1192. <https://doi.org/10.1016/j.jclepro.2017.09.098>.
 [6] A.C. Emmanuel, P. Haldar, S. Maity, S. Bishnoi, Second pilot production of limestone calcined clay cement in India: The experience, *Indian Concr. J.* 90 (2016) 57–63.
 [7] S. Bishnoi, S. Maity, A. Mallik, S. Joseph, S. Krishnan, Pilot scale manufacture of limestone calcined clay cement : The Indian experience, *Indian Concr. J.* 88 (2014) 22–28.
 [8] S.M. Gupta, Support Vector Machines based Modelling of Concrete Strength, *World Acad. Sci. Eng. Technol.* 36 (2007) 305–311.
 [9] S. Lai, M. Serra, Concrete strength prediction by means of neural network, *Constr. Build. Mater.* 11 (1997) 93–98. [https://doi.org/10.1016/S0950-0618\(97\)00007-X](https://doi.org/10.1016/S0950-0618(97)00007-X).
 [10] M. Sarıdemir, Prediction of compressive strength of concretes containing metakaolin and silica fume by artificial neural networks, *Adv. Eng. Softw.* 40 (2009) 350–355. <https://doi.org/10.1016/j.advengsoft.2008.05.002>.
 [11] A. Chahal, P. Gulia, Machine learning and deep learning, *Int. J. Innov. Technol. Explor. Eng.* 8 (2019) 4910–4914. <https://doi.org/10.35940/ijitee.L3550.1081219>.
 [12] C. Deepa, K. Sathiyakumari, V.P. Sudha, Prediction of the Compressive Strength of High Performance Concrete Mix using Tree Based Modeling, *Int. J. Comput. Appl.* 6 (2010) 18–24.

- با توجه به شکل ۱ مواد جایگزین سیمانی در هر مدل تأثیر تقریباً یکسانی بر روی مقاومت فشاری داشته اما رس با توجه به ویژگی‌های آن دارای مزیت‌های بیشتری در مقایسه با بقیه مواد جایگزین سیمانی است. همچنین استفاده از رس کلسینه به میزان بهینه در بلند مدت باعث افزایش مقاومت فشاری بتن شده که این مقدار اگر فراتر رود باعث کاهش آن می‌شود.

- با توجه به صحت سنجی انجام شده، مدل‌های یادگیری شده قابلیت پیش‌بینی مقاومت فشاری داده‌هایی که مدل ندیده است را دارا می‌باشند به شرطی که مدل توزیع مشابهی از این داده‌ها دیده باشد. همچنین می‌توان نتیجه گرفت که بیش برآزش در مدل‌های XGBoost، LightGBM و جنگل تصادفی اتفاق نیافتاده که نشان دهنده عملکرد مناسب این سه مدل و سازگاری با الگوریتم ژنتیک است.

۶- مراجع

[1] K. Scrivener, Options for the future of cement, *Indian Concr. J.* Vol. 88 (2014) 11–21.
 [2] A. K. Scrivener, A. Dekeukelaere, L.G. F. Avet, Financial Attractiveness of LC3, (2019).
 [3] A. Jain, A. Fandango, A. Kapoor, TensorFlow Machine Learning Projects: Build 13 Real-World Projects with Advanced Numerical Computations Using the Python Ecosystem, Packt Publishing, 2018.
 [4] K. Scrivener, F. Martirena, S. Bishnoi, S. Maity, Calcined clay limestone cements (LC3), *Cem. Concr. Res.* 114 (2018) 49–56.

- <https://doi.org/10.4028/www.scientific.net/AMM.567.375>.
- [24] L. Vizcaíno, M. Antoni, A. Alujas, F. Martirena, K. Scrivener, Industrial Manufacture of a Low-Clinker Blended Cement Using Low-Grade Calcined Clays and Limestone as SCM: The Cuban Experience, in: RILEM Bookseries, 2015: pp. 347–358. https://doi.org/10.1007/978-94-017-9939-3_43.
- [25] I.-C. Yeh, Modeling slump of concrete with fly ash and superplasticizer, *Comput. Concr.* 5 (2008) 559–572. <https://doi.org/10.12989/cac.2008.5.6.559>.
- [26] A. SAAND, M.A. KEERIO, D. khan BANGWAR, EFFECT OF METAKAOLIN DEVELOPED FROM LOCAL NATURAL MATERIAL SOORH ON WORKABILITY, COMPRESSIVE STRENGTH, ULTRASONIC PULSE VELOCITY AND DRYING SHRINKAGE OF CONCRETE, *Archit. Civ. Eng. Environ.* 10 (2017) 115–122. <https://doi.org/10.21307/acee-2017-025>.
- [27] H.A. Razak, H.S. Wong, Strength estimation model for high-strength concrete incorporating metakaolin and silica fume, *Cem. Concr. Res.* 35 (2005) 688–695. <https://doi.org/10.1016/j.cemconres.2004.05.040>.
- [28] K. Shui, K. Yuan, T. Sun, Q. Li, W. Zeng, *Calcined Clays for Sustainable Concrete*, Springer Netherlands, Dordrecht, 2015. <https://doi.org/10.1007/978-94-017-9939-3>.
- [29] K. Mermerdaş, M. Gesođ lu, E. Güneyisi, T. Özturan, Strength development of concretes incorporated with metakaolin and different types of calcined kaolins, *Constr. Build. Mater.* 37 (2012) 766–774. <https://doi.org/10.1016/j.conbuildmat.2012.07.077>.
- [30] R.R. Raj, E.B.P. Pillai, Shear Strength of High Performance Concrete Containing High Reactivity Metakaolin under Direct Shearing, *Int. J. Environ. Sci. Technol.* (2008).
- [31] S.A. Zareei, F. Ameri, F. Dorostkar, M. Ahmadi, Rice husk ash as a partial replacement of cement in high strength concrete containing micro silica: Evaluating durability and mechanical properties, *Case Stud. Constr. Mater.* 7 (2017) 73–81. <https://doi.org/10.1016/j.cscm.2017.05.001>.
- [32] A.F. Karen Scrivener, Ruben Snellings, Xuerun Li, *Cement Chemistry and Sustainable Cementitious Materials course*, (n.d.). <https://www.edx.org/course/cement-chemistry-and-sustainable-cementitious-mate>.
- [33] American Society for Testing and Materials, ASTM C 150 : Standard Specification for Portland Cement, *Annu. B. ASTM Stand.* 04.01 (2001) 149–<https://doi.org/10.5120/1076-1406>.
- [13] S. Dutta, P. Samui, D. Kim, Comparison of machine learning techniques to predict compressive strength of concrete, *Comput. Concr.* 21 (2018) 463–470. <https://doi.org/10.12989/cac.2018.21.4.463>.
- [14] J.-S. Chou, C.-K. Chiu, M. Farfoura, I. Al-Taharwa, Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques, *J. Comput. Civ. Eng.* 25 (2011) 242–253. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000088](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000088).
- [15] S. Iii, *Machine Learning Methods for Predicting the Field Compressive Strength of Concrete*, (2019).
- [16] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, *Neurocomputing.* 415 (2020) 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>.
- [17] L.O. Ettu, C.A. Ajoku, K.C. Nwachukwu, C.T.G. Awodiji, U.G. Eziefula, Strength variation of OPC-rice husk ash composites with percentage rice husk ash, *Int. J. Appl. Sci. Eng. Res.* 2 (2013) 420–424. <https://doi.org/10.6088/ijaser/020400004>.
- [18] E. Opoku Amankwah, Influence of Calcined Clay Pozzolana on Strength Characteristics of Portland Cement Concrete, *Int. J. Mater. Sci. Appl.* 3 (2014) 410. <https://doi.org/10.11648/j.ijmsa.20140306.30>.
- [19] R.M. Ferreira, J.P. Castro-Gomes, P. Costa, R. Malheiro, Effect of metakaolin on the chloride ingress properties of concrete, *KSCE J. Civ. Eng.* 20 (2016) 1375–1384. <https://doi.org/10.1007/s12205-015-0131-8>.
- [20] G. Kaplan, H. Yaprak, S. Memiş, A. Alnkaa, Artificial neural network estimation of the effect of varying curing conditions and cement type on hardened concrete properties, *Buildings.* 9 (2019). <https://doi.org/10.3390/buildings9010010>.
- [21] D. Vu, P. Stroeven, V. Bui, Strength and durability aspects of calcined kaolin-blended Portland cement mortar and concrete, *Cem. Concr. Compos.* 23 (2001) 471–478. [https://doi.org/10.1016/S0958-9465\(00\)00091-3](https://doi.org/10.1016/S0958-9465(00)00091-3).
- [22] R.-S. Lin, X.-Y. Wang, H.-S. Lee, H.-K. Cho, Hydration and Microstructure of Cement Pastes with Calcined Hwangtoh Clay, *Materials (Basel).* 12 (2019) 458. <https://doi.org/10.3390/ma12030458>.
- [23] M.F. Nuruddin, S.U. Khan, N. Shafiq, Effect of Calcined Kaolin on the Mechanical Properties of High-Strength Concrete as Cement Replacing Material, *Appl. Mech. Mater.* 567 (2014) 375–380.

- [45] Y. Freund, R.E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *J. Comput. Syst. Sci.* 55 (1997) 119–139. <https://doi.org/10.1006/jcss.1997.1504>.
- [46] V. Chandwani, V. Agrawal, R. Nagar, Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks, *Expert Syst. Appl.* 42 (2015) 885–893. <https://doi.org/10.1016/j.eswa.2014.08.048>.
- [34] M. Bediako, J.T. Kevern, J.S. Ankrah, Strength and durability of cement-based materials incorporated with low grade kaolinitic calcined clay, *Sustain. Constr. Mater. Technol.* 2016-Augus (2016).
- [35] I.-C.C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, *Cem. Concr. Res.* 28 (1998) 1797–1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3).
- [36] D. Zhou, Developing Supplementary Cementitious Materials From Waste London Clay, (2016) 236. <https://spiral.imperial.ac.uk/handle/10044/1/44528>.
- [37] M. Sullivan, M. Chorzepa, H. Hamid, S. Durham, S. Kim, Sustainable Materials for Transportation Infrastructures: Comparison of Three Commercially-Available Metakaolin Products in Binary Cementitious Systems, *Infrastructures.* 3 (2018) 17. <https://doi.org/10.3390/infrastructures3030017>.
- [38] K. Ganesan, K. Rajagopal, K. Thangavel, Rice husk ash blended cement: Assessment of optimal level of replacement for strength and permeability properties of concrete, *Constr. Build. Mater.* 22 (2008) 1675–1683. <https://doi.org/10.1016/j.conbuildmat.2007.06.011>.
- [39] I.-C. Yeh, L.-C. Lien, Knowledge discovery of concrete material using Genetic Operation Trees, *Expert Syst. Appl.* 36 (2009) 5807–5812. <https://doi.org/10.1016/j.eswa.2008.07.004>.
- [40] H.I. Erdal, O. Karakurt, E. Namli, High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform, *Eng. Appl. Artif. Intell.* 26 (2013) 1246–1254. <https://doi.org/10.1016/j.engappai.2012.10.014>.
- [41] D.-C. Feng, Z.-T. Liu, X.-D. Wang, Y. Chen, J.-Q. Chang, D.-F. Wei, Z.-M. Jiang, Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach, *Constr. Build. Mater.* 230 (2020) 117000. <https://doi.org/10.1016/j.conbuildmat.2019.117000>.
- [42] S. Zhang, C. Zhang, Q. Yang, Data preparation for data mining, *Appl. Artif. Intell.* 17 (2003) 375–381. <https://doi.org/10.1080/713827180>.
- [43] J.F. Martirena Hernandez, A. Favier, K. Scrivener, *Calcined Clays for Sustainable Concrete*, Springer Netherlands, Dordrecht, 2018. <https://doi.org/10.1007/978-94-024-1207-9>.
- [44] E. Al Daoud, Comparison between XGBoost , LightGBM and CatBoost Using a Home Credit Dataset, 13 (2019) 6–10. <https://doi.org/10.5281/zenodo.3607805>.

Data Mining Mechanical properties of concretes containing calcined clays as supplementary cementitious materials in concrete

Ali Hosein Ghanemi *

Master of Science, Department of Civil Engineering, Shahid Rajaei Teacher Training University.

Amir Tarighat

Associate Professor, Department of Civil Engineering, Shahid Rajaei Teacher Training University.

Abstract

Concrete due to its excellent properties, low cost and wide availability is the most used construction product in the world. As the world population grows, cement production has increased as one of the main materials used in concrete. Cement production emits carbon dioxide which increases environmental pollution. One of the ways to prevent pollution is to use Supplementary cementing materials instead of cement. One of the most important alternative materials used in recent years is LC3 cement. This type of cement reduces the amount of clinker in cement, decreases our need in fossil fuels and therefore reduces the emission of carbon dioxide. Moreover, one of the most important mechanical properties of concrete is its compressive strength, which its estimation is complex since the number of existing parameters is high. Hence, costly laboratory methods are used which consist of high error. In this paper, machine learning models that work based on decision tree model are used which their performance of have been improved by genetic algorithm. LightGBM and XGBoost model got a prediction score of 0.958 in prediction the compressive strength of concrete and perform better than decision tree and random forest models with 0.91 and 0.932 prediction score. Also, feature importance of each model have been presented and a new data set have been used to evaluate the validation of models.

Keywords: Concrete, Machine Learning and Data Mining, LC3 cement, Supplementary cement materials, Boosting algorithm.

* Corresponding Author: shayan.ghanemi@sru.ac.ir