19th Iranian Soil Science Congress
2-4 December, 2025

نوزدهمین کنگره علوم خاک ایران
۱۱ تا ۱۳ آذرماه ۱۴۰۴

مدیریت جامع‌نگر و هوشمند خاک و آب
Holistic and Smart Soil and Water Management
دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران
College of Agriculture & Natural Resources, University of Tehran

۰۴۲۵۰-۳۲۰۳۱

تخمین مقدار ظرفیت تبادل کاتیونی (CEC) بر پایه ویژگی‌های زودیافت خاک با استفاده از الگوریتم جنگل تصادفی

سمیه شامی هریس[۱*]، محبوبه مظهری[۲]، مهناز اسکندری[۳]

[۱]دانشجوی دکتری منابع خاک و ارزیابی اراضی، دانشگاه گیلان، (shamishami1362@gmail.com)
[۲]استاد گروه علوم و مهندسی خاک، دانشگاه آزاد اسلامی واحد کرج، ایران
[۳]استادیار پژوهش، موسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران

چکیده:

ظرفیت تبادل کاتیونی یکی از شاخص‌های مهم برای تعیین کیفیت و بهره‌وری خاک است. از آنجا که اندازه‌گیری این ویژگی، پرهزینه، زمان‌بر و نیازمند کاربرد دستگاه‌های آزمایشگاهی ویژه‌ای است، تخمین آن به کمک روش‌های اندازه‌گیری غیر مستقیم می‌تواند مقرون‌به‌صرفه باشد. هدف از انجام این پژوهش، پیش‌بینی ظرفیت تبادل کاتیونی خاک به کمک برخی ویژگی‌های زودیافت خاک با استفاده از سه مدل جنگل تصادفی، درخت تصمیم و رگرسیون خطی بود. نخست، ۱۸۸ نمونه مرکب خاک به روش نمونه‌برداری منظم با فاصله یک کیلومتر، از دشت داراب استان فارس تهیه شد. سپس ظرفیت تبادل کاتیونی ($Cmol^{+}kg^{-1}$)، Hp، درصد آهک، کربن آلی، درصد رس، شن و سیلت در نمونه‌ها اندازه‌گیری شد. با کمک ۷۰ درصد داده‌ها، مدل‌های مورد نظر طراحی و با ۳۰ درصد باقیمانده، اعتبارسنجی یا آزمون مدل انجام پذیرفت. نتایج همبستگی نشان داد ظرفیت تبادل کاتیونی دارای همبستگی مثبت و معنی‌دار با مقدار رس خاک (۰/۷۳) و کربن آلی (۰/۴۶) و همبستگی منفی و معنی‌دار با مقدار شن (۰/۵۶-) است. مقایسه نتایج حاصل از سه مدل نشان داد که جنگل تصادفی به ترتیب با ضریب تبیین ۰/۶۰ و RMSE برابر ۲/۶۵۸، مناسب‌ترین مدل برای برآورد ظرفیت تبادل کاتیونی می‌باشد.

واژگان کلیدی: CEC، الگوریتم جنگل تصادفی، درخت تصمیم، رگرسیون خطی

مقدمه:

خاک یک منبع حیاتی است که خدمات اکوسیستمی متنوعی مانند تولید غذا، چرخه کربن و مواد مغذی ارائه می‌دهد. ازاین‌رو حفظ خاک امری حیاتی و ضروری است تا برای حمایت از تنوع زیستی برای زندگی انسان و شرایط اجتماعی-اقتصادی، مورد توجه قرار گیرد (ماریو پالادینو و همکاران، ۲۰۲۲). ظرفیت تبادل کاتیونی (CEC) [۱] شاخص مهمی برای تعیین کیفیت و بهره‌وری خاک است. این ویژگی بر پایداری ساختاری خاک، در دسترس بودن مواد مغذی، pH و واکنش به کودها تأثیرگذار است (زوئیو ژائو و همکاران، ۲۰۲۰). ظرفیت تبادل کاتیونی معمولاً با واحد میلی اکی والان در ۱۰۰ گرم خاک (meq 100g-1) یا سانتی مول بار بر کیلوگرم خاک (cmol (+) kg-1) بیان می‌شود (جمشیدی، ۱۳۹۱).

با توجه به اهمیت ظرفیت تبادل کاتیونی، تاکنون مطالعات متعددی در زمینه مقایسه کارایی روش‌های تخمین مانند فازی[۲]،

[1] Cation Exchange Capacity (CEC)
[2] Fuzzy logic

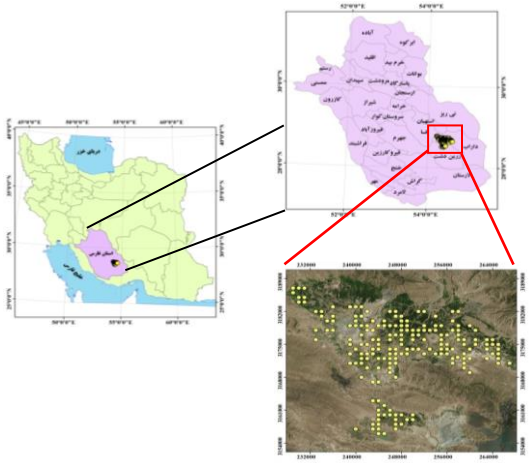
شبکه‌های عصبی مصنوعی[3] و نروفازی[4] انجام شده است (کدخدایی و همکاران، ۲۰۰۹). همچنین، ماشین‌های یادگیری[5] نیز روش‌های جدید برای پیش‌بینی ویژگی‌های خاک می‌باشند. درخت تصمیم[6] یک ابزار برای پشتیبانی از تصمیم است که به‌طور معمول در تحقیق‌ها و عملیات مختلف استفاده می‌شود.

الگوریتم جنگل تصادفی (RF)[7] یکی از انواع روش‌های یادگیری و داده‌کاوی است. این روش در سال ۲۰۰۱ توسط بریمن ارائه شد. یکی از الگوریتم‌های مورد استفاده در این پژوهش مدل جنگل تصادفی است. این الگوریتم به دلیل سادگی و قابلیت استفاده، هم برای دسته‌بندی (Classification) و هم رگرسیون (Regression)، یکی از پرکاربردترین الگوریتم‌های یادگیری ماشین محسوب می‌شود. در این روش چندین درخت تصمیم آموزش داده می‌شود تا دقت مدل را افزایش دهند. بنابراین نتیجه حاصل پیش‌بینی گروهی از درختان تصمیم‌گیری است (بریمان، ۲۰۰۱).

خاموشی و همکاران (۱۳۹۷) برای تهیه نقشه رقومی‌خاک در منطقه آبیک استان قزوین از مدل جنگل تصادفی استفاده کردند. در این پژوهش به روش مکعب لاتین مشروط از منطقه نمونه‌برداری شد و پس از انجام آنالیزهای آزمایشگاهی، با استفاده از مدل جنگل‌های تصادفی و متغیرهای کمکی مستخرج از مدل رقومی‌ارتفاعی با قدرت تفکیک مکانی ۳۰ متر و تصاویر ماهواره لندست ۸، نقشه خاک منطقه تهیه شد. نتایج حاصل نشان داد که مدل توانسته است کلاس‌های خاک منطقه را با دقت بالایی (ضریب کاپای ۰/۸۳) پیش‌بینی کند. هدف از انجام این پژوهش ۱) تخمین ظرفیت تبادل کاتیونی خاک با استفاده از برخی خصوصیات فیزیکی و شیمیایی زودیافت خاک، ۲) تعیین مهم‌ترین ویژگی‌های مؤثر در تخمین ظرفیت تبادل کاتیونی در خاک‌های منطقه انتخابی و ۳) مقایسه دقت مدل جنگل تصادفی (RF) با رگرسیونی خطی و درخت تصمیم از طریق ارزیابی عملکرد آن‌ها در خاک‌های منطقه انتخابی بود.

مواد و روش‌ها :

منطقه مورد مطالعه در استان فارس در غرب و جنوب شهر داراب در مختصات جغرافیایی ″۳۶ ′۲۳ °۵۴ تا ″۳۶ ′۳۵ °۵۴ طول شرقی و ″۵۷ ′۳۷ °۲۸ تا ″۰۲ ′۵۰ °۲۸ عرض شمالی قرار گرفته است. شکل ۱ موقعیت منطقه مورد مطالعه را در استان فارس نشان می دهد.



[3] Artificial Neural Networks
[4] Neurofuzzy
[5] Machine learning
[6] Decision Tree
[7] Random Forest

شکل ۱- موقعیت منطقه مورد مطالعه و محل نقاط نمونه‌برداری شده در استان فارس

محدوده مورد مطالعه داراب بر اساس طبقه‌بندی جامع خاک‌ها (Soil Taxonomy) در رده‌های انتی سول (Entisols)، اینسپتی سول (Inceptisols) و ورتی سول (Vertisols) در پنج گروه بزرگ و در شش زیرگروه و ۱۴ فامیل طبقه‌بندی گردیده است. متوسط بارندگی سالانه در منطقه در حدود ۲۵۷ میلی‌متر و ماکزیمم سالانه درجه حرارت آن ۴۰ درجه و مینیمم سالانه دما در منطقه ۱۴/۴ درجه است. رژیم رطوبتی، Ardic Ustic و حرارتی، Hyperthermic می‌باشد (جمشیدی، شرکت مهندسی مشاور حاسب کرجی، ۱۳۹۱).

داده‌های مورد استفاده در این پژوهش، از مطالعه " گزارش پهنه‌بندی ظرفیت تبادل کاتیونی خاک‌ها به منظور بررسی توان خاک‌ها در نگهداری و انتقال عناصر آلاینده در استان‌های خوزستان و فارس " مربوط به استان فارس که در سال ۱۳۹۱ توسط محققین موسسه تحقیقات خاک و آب انجام شده است، برداشت شد. مساحت نمونه‌برداری در منطقه مورد نظر در حدود ۵۳۲۱۷ هکتار بود که ۴۷۸۰۰ هکتار آن در دشت داراب و ۵۴۱۷ هکتار آن در زرین دشت واقع شده است. ۱۸۸ نمونه مرکب خاک از عمق ۳۰-۰ سانتی‌متری به روش نمونه‌برداری منظم با فاصله یک کیلومتر از یکدیگر در این منطقه برداشت شده است. سپس ویژگی‌های خاک برای هر نمونه در آزمایشگاه اندازه‌گیری شد. در این مطالعه CEC خاک‌ها به روش باور (رودز، ۱۹۸۲)، بافت خاک به روش هیدرومتری (Bouyoucos, 1962)، کربن آلی (OC) به روش روش واکلی و بلاک (Walkley and Black, 1934)، ظرفیت تبادل کاتیونی (CEC) به روش جانشینی کاتیون‌ها با سدیم استات (Chapman, 1965)، واکنش خاک (pH) به روش pHمتر و آهک (TNV) به روش تیتراسیون اندازه‌گیری شد (بی‌نام، ۱۳۸۷).

مدل‌های رگرسیون خطی[8]، درخت تصمیم[9] و الگوریتم جنگل تصادفی[10] برای تخمین مقدار ظرفیت تبادل کاتیونی (CEC) به کمک داده‌های زودیافت خاک استفاده شد. در این پژوهش از متغیرهای کمکی مؤثر در ظرفیت تبادل کاتیونی استفاده شد که با استفاده از انتقال متغیرها در قالب یک فایل عددی با فرمت (CSV) از بسته 'Random Forest' به محیط نرم‌افزار R Studio نسخه 1.1.456 به همراه کد نویسی صورت پذیرفت. نخست باید مجموعه داده‌های به‌دست‌آمده را به دو دسته آموزش و آزمون طبقه‌بندی نمود. بنابراین با استفاده از ۷۰ درصد داده‌ها مدل طراحی شد و با ۳۰ درصد باقیمانده، اعتبار سنجی یا آزمون مدل انجام پذیرفت. هر مجموعه داده شامل داده‌های ورودی یعنی کربن آلی، واکنش خاک، آهک و بافت و مقدار خروجی یعنی ظرفیت تبادل کاتیونی است. جنگل تصادفی به‌عنوان یک مدل یادگیری ماشین نظارت‌شده، یاد می‌گیرد که در فاز آموزش یا برازش مدل، داده‌های ورودی را به خروجی نگاشت کند. پس از اتمام آموزش و به دست آوردن آماره‌هایی مانند R2 و RMSE که نشان از صحت کارکرد مدل داشته باشه، داده‌های آزمون به مدل وارد می‌شوند. پس از محاسبه مقدار خروجی برای داده‌های آزمون، مقادیر پیش‌بینی‌شده با مقادیر واقعی داده‌ها مقایسه شد تا دقت کارکرد مدل طراحی‌شده مشخص گردد.

ارزیابی مدل‌ها برپایه برخی آماره‌ها شامل حداکثر خطا (ME)، میانگین ریشه دوم خطا (RMSE)، ضریب تبیین (CD)، کارآیی مدل‌سازی (EF)، و ضریب باقیمانده (CRM) انجام شد (فلامکی و اسکندری، ۱۳۹۱).

رابطه (۱)
$$ME = \max\left|P_i - O_i\right|_{i=1}^{n}$$

رابطه (۲)
$$RMSE = \left[\frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}\right]^{0.5} \frac{100}{\overline{O}}$$

[8] Simple Linear Regression
[9] Decision Tree
[10] Random Forest

رابطه (۳)

$$CD = \frac{\sum_{i=1}^{n}(O_i - \overline{O})^2}{\sum_{i=1}^{n}(p_i - \overline{O})^2}$$

رابطه (۴)

$$EF = \frac{\sum_{i=1}^{n}(O_i - \overline{O})^2 - \sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$$

رابطه (۵)

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i}$$

در رابطه‌ها Pi مقدار پیش‌بینی‌شده، Oi مقدار اندازه‌گیری شده (مشاهده‌شده)، n تعداد نمونه‌های به کار رفته و $\overline{O}$ مقدار متوسط مقادیر O است. حداکثر مقدار EF برابر یک است. حداقل مقدار ME، RMSE و CD صفر است. مقدار منفی EF بیانگر آن است که میانگین مقادیر اندازه‌گیری شده، برآوردی بهتر از مقادیر پیش‌بینی‌شده دارد. آماره CRM نشانگر تمایل مدل برای بیش برآوردی و یا کم‌برآوردی در مقایسه با اندازه‌گیری‌ها است. چنانچه تمام مقادیر اندازه‌گیری شده و برآورد شده با هم برابر باشند، مقدار عددی این آماره‌ها برابر است با: ME=۰، RMSE=۰، CD=۱، EF=۱ و CRM=۰ (فلامکی و اسکندری، ۱۳۹۱).

نتایج و بحث :

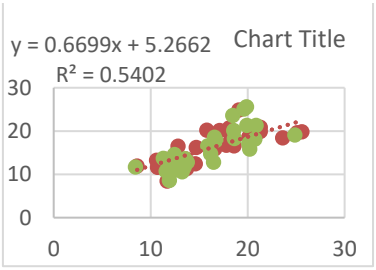
نتایج مربوط به آمار توصیفی ویژگی‌های مورد بررسی خاک در این پژوهش، شامل میانگین، کمینه، بیشینه، انحراف معیار، واریانس، چولگی و کشیدگی، در جدول ۱ ارائه شده است. با توجه به نتایج به دست آمده کمترین و بیشترین مقدار ظرفیت تبادل کاتیونی به ترتیب ۵/۱ $Cmol^+ kg^{-1}$ و ۲۸/۶ $Cmol^+ kg^{-1}$ و با میانگین ۱۶/۴۴۴ $Cmol^+ kg^{-1}$ بدست آمد. همچنین مقدار انحراف معیار نمونه خاک‌های مربوطه برای ظرفیت تبادل کاتیونی برابر ۴/۴۸۴۷ می‌باشد. حداقل و حداکثر مقدار pH خاک برابر ۷/۱۷ و ۸/۶۰ با میانگین ۷/۹۱ است که نشان می‌دهد خاک‌های مورد مطالعه در محدوده خنثی تا قلیایی قرار دارند. پس از بررسی درصد آهک (TNV) در ۱۸۸ نمونه مورد بررسی نشان داده شد که تنوع صفت آهک در خاک‌های مورد بررسی زیاد بوده به طوریکه کمترین مقدار آهک برابر ۲۱/۵ و ۷۴/۵ درصد با میانگین ۴۵/۰۱۲ درصد بدست آمد. مقدار درصد کربن آلی در نمونه خاک‌های مورد مطالعه دارای میانگین ۰/۹۶ با حداقل و حداکثر مقدار ۰/۱۷ و ۲/۵۸ درصد بود. خاک‌های مورد بررسی دارای مقادیر متنوع شن، سیلت و رس بودند به طوریکه مقدار کمینه و بیشینه برای شن به ترتیب برابر ۵/۲ و ۷۷/۴، برای سیلت ۱۱/۰ و ۶۳/۲ و برای رس برابر ۱۰/۰ و ۵۴/۰ درصد بود.

جدول ۱- آماره‌های توصیفی برای ۱۸۸ نمونه خاک مورد بررسی در منطقه مورد مطالعاتی

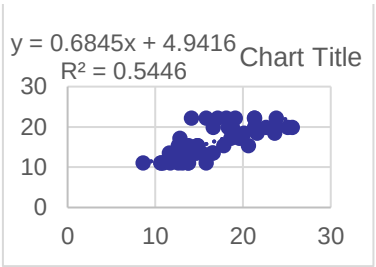
خصوصیات	تعداد نمونه	کمینه	بیشینه	میانگین	انحراف معیار	واریانس	چولگی		کشیدگی	
ظرفیت تبادل کاتیونی (CEC)$Cmol^+ kg^{-1}$	۱۸۸	۱/۵	۶/۲۸	۴۴۴/۱۶	۴۸۴۷/۴	۱۱۳/۲۰	۴۶۰/۰	۱۷۷/۰	۲۸۲/۰-	۳۵۳/۰
واکنش خاک (pH)	۱۸۸	۱۷/۷	۶۰/۸	۹۱۴۰/۷	۲۲۹۰۱/۰	۰۵۲/۰	۱۶۹/۰	۱۷۷/۰	۸۳۶/۰	۳۵۳/۰
آهک TNV (%)	۱۸۸	۲۱/۵	۷۴/۵	۴۵/۰۱۲	۶/۶۷۶۶	۴۴/۵۷۸	۰/۹۸۴	۰/۱۷۷	۳/۳۷۳	۰/۳۵۳
کربن آلی (%)	۱۸۸	۰/۱۷	۲/۵۸	۰/۹۶۳۲	۰/۳۷۳۲۴	۰/۱۳۹	۰/۵۷۸	۰/۱۷۷	۱/۴۰۳	۰/۳۵۳
شن SAND%	۱۸۸	۵/۲	۷۷/۴	۲۷/۲۱۰	۱۴/۵۷۲۴	۲۱۲/۳۵۴	۰/۹۸۰	۰/۱۷۷	۰/۸۳۴	۰/۳۵۳
سیلت SILT%	۱۸۸	۱۱/۰	۶۳/۲	۴۱/۲۱۳	۹/۵۳۴۵	۹۰/۹۰۷	-۰/۴۱۹	۰/۱۷۷	۰۰/۵۵۵	۰/۳۵۳
رس CLAY %	۱۸۸	۱۰/۰	۵۴/۰	۳۱/۵۷۷	۹/۶۳۵۱	۹۲/۸۳۵	-۰/۰۴۳	۰/۱۷۷	-۰/۴۶۱	۰/۳۵۳

الگوی پراکنش نقاط بر روی مثلث بافت خاک نشان‌دهنده تجمع نقاط در قسمت مرکزی است. بنابراین بیشتر نقاط در بافت‌های لوم، کلی لوم و سیلتی کلی لوم قرار گرفته‌اند. هیچ یک از نقاط نمونه‌برداری شده در کلاس‌های بافتی sand, loamy sand, silt, sandy clay قرار نگرفته است. از طرفی کمترین تعداد داده‌ها در کلاس بافتی sandy clay loam و بیشترین در کلاس بافتی clay loam می‌باشد.

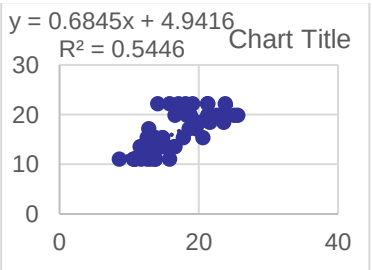
نتایج همبستگی پیرسون بین متغیرهای خاک در ۱۸۸ نمونه مورد بررسی در منطقه مطالعاتی نشان داد که ظرفیت تبادل کاتیونی دارای همبستگی مثبت و معنی‌دار با مقدار رس خاک (۰/۷۳) و کربن آلی (۰/۴۶) می‌باشد. بیشترین همبستگی منفی و معنی‌دار را با مقدار شن (۰/۵۶-) داشت. بنابراین می‌توان نتیجه گرفت همبستگی بالای ظرفیت تبادل کاتیونی با درصد رس و کربن آلی ناشی از این است که این متغیر به شدت تحت تاثیر این دو صفت قرار دارد و می‌توانند در پیش‌بینی ظرفیت تبادل کاتیونی نقش اساسی داشته باشند. نتایج حاصل از پیش‌بینی مقدار ظرفیت تبادل کاتیونی با استفاده از رگرسیون خطی و درخت تصمیم به ترتیب ۵۴/۰۲ و ۵۴/۴۶ درصد را نشان می دهد (شکل۲ و ۳). همچنین مقادیر پیش‌بینی شده ظرفیت تبادل کاتیونی توسط مدل جنگل تصادفی در شکل ۵ نیز ارائه شده است.



شکل ۲- مقادیر پیش‌بینی ظرفیت تبادل کاتیونی توسط مدل رگرسیون خطی



شکل ۳- مقادیر پیش‌بینی شده ظرفیت تبادل کاتیونی توسط مدل درخت تصمیم

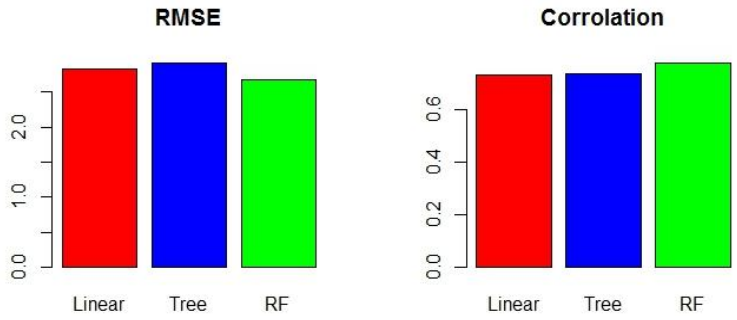


شکل ۴- مقادیر پیش‌بینی شده ظرفیت تبادل کاتیونی توسط مدل جنگل تصادفی

براساس نتایج به دست آمده از روش جنگل تصادفی به ترتیب صفات درصد رس، درصد شن، کربن آلی، درصد آهک و pH بیشترین تاثیر را در مدل داشته‌اند. مقایسه سه مدل جنگل تصادفی، درخت تصمیم و رگرسیون خطی در جدول ۲ ارائه شده است که نشان می دهدبیشترین مقدار ضریب تبیین مربوط به مدل جنگل تصادفی با مقدار ۰/۶۰ درصد و به ترتیب بعد از آن مدل رگرسیون خطی و درخت تصمیم با ضریب تعیین ۰/۵۴ و ۰/۵۴ درصد است. نتایج ارزیابی مدل‌های مورد بررسی در پیش‌بینی ظرفیت تبادل کاتیونی براساس R^2 و RMSE در شکل ۵ ارائه شده است.

جدول ۲- آماره‌های ارزیابی مدل‌های مورد بررسی در پیش‌بینی ظرفیت تبادل کاتیونی

	CD	RMSE	EF	MAE	CRM
رگرسیون خطی	۵۴۰/۰	۲/۸۲۷	۰/۵۳۶	۲/۳۰۵	۰/۰۰۱
درخت تصمیم	۵۴۴/۰	۲/۹۱۰	۰/۵۰۸	۲/۲۹۵	۰/۰۱۲۳
جنگل تصادفی	۶۰۹/۰	۲/۶۵۸	۰/۵۹۰	۲/۱۵۶	-۰/۰۰۱



شکل ۵- ارزیابی صحت مدل‌های مورد بررسی در پیش‌بینی ظرفیت تبادل کاتیونی براساس R و RMSE

هرچه مقدار میانگین ریشه دوم خطا (RMSE) کمتر باشد نشان دهنده دقت بالای مدل است. بر اساس نتایج، مقدار RMSE در مدل‌های جنگل تصادفی، رگرسیون خطی و درخت تصمیم به ترتیب برابر است با ۲/۶۵۸، ۲/۸۲۷ و ۲/۹۱۰ که نشان می‌دهد مدل جنگل تصادفی بهترین مدل برای برآورد ظرفیت تبادل کاتیونی است. براساس کارایی مدل (EF) نیز مدل جنگل تصادفی با مقدار ۰/۵۹۰ دارای بالاترین کارایی است. هرچه مقدار CRM به صفر نزدیک باشد نشان می‌دهد که مدل برآورد مناسبی از مقادیر CEC داشته است.

نتیجه‌گیری:

مقایسه سه مدل جنگل تصادفی، درخت تصمیم و رگرسیون خطی در برآورد مقدار CEC خاک نشان داد که مدل جنگل تصادفی با مقدار ضریب تبیین ۶۰/۸۴ و بعد از آن مدل درخت تصمیم و رگرسیون خطی با ضریب تبیین ۵۴/۴۶ و ۵۴/۰۲ به ترتیب بهترین مدل می‌باشند. همچنین بر اساس نتایج به دست آمده مدل جنگل تصادفی دارای کمترین مقدار RMSE و بالاترین کارایی مدل (EF) بود. بنابراین می‌توان چنین نتیجه گرفت که جنگل تصادفی یک روش آسان و مقرون به صرفه برای پیش‌بینی ظرفیت تبادل کاتیونی خاک است که می‌تواند به تخمین مقدار آن به‌جای اندازه‌گیری این ویژگی کمک نماید.

فهرست منابع:

۱. بی‌نام. ۱۳۸۷. دستورالعمل تجزیه‌های آزمایشگاهی نمونه‌های خاک و آب، نشریه شماره ۴۶۷. معاونت برنامه‌ریزی و نظارت راهبردی رئیس‌جمهور.
۲. جمشیدی، م. ۱۳۹۱. گزارش پهنه‌بندی ظرفیت تبادل کاتیونی خاک‌ها به منظور بررسی توان خاک‌ها در نگهداری و انتقال عناصر آلاینده در استان‌های خوزستان و فارس، شرکت مهندسی مشاور حاسب کرجی، ۱۳۹۱).
۳. فلامکی، ا. و اسکندری، م. ۱۳۹۱. گزارش تخمین ضریب توزیع خاک-آب فلزات سنگین با کاربرد شبکه های عصبی مصنوعی، نشریه علمی-پژوهشی حفاظت منابع آب و خاک.
4. Breiman, L., Random Forests. Machine Learning. 2001; 45(1): 5–32.
5. Kadkhodaie, A., Rezaee, MR, Rahimpour-Bonab, H. A. committee neural network for prediction of normalized oil content from well log data: An example from South Pars GasField, Persian Gulf. Journal of Petroleum Science and Engineering. 2009; 65: 23-32.

Estimating the amount of cation exchange capacity based on the characteristics of soil precipitation with the help of some transfer functions

somayeh shami[*1], Mehbobe Mazhari[2], Mahnaz Eskandari[3]

[1] PhD student of soil resources and land evaluation, Gilan University,
[2] Professor, Department of Soil Science and Engineering, Islamic Azad University, Karaj Branch, Iran
[3] Research Assistant Professor, Soil and Water Research Institute, Agricultural Research, Education and Extension Organization, Karaj, Iran
[*]shamishami1362@gmail.com

Abstract

Cation exchange capacity (CEC) is one of the important indicators for determining soil quality and productivity. Since the measurement of this characteristic is costly, time-consuming and requires special laboratory devices, Therefore, it is estimated by using soil transfer. In order to predict the Cation exchange capacity of soil using soil characteristics, soil samples were prepared from 188 samples from Darab region of Fars province and the characteristics of CEC (Cmol+ kg-1), pH, percentage of lime (TNV), organic carbon (OC) and the percentage of clay, sand and silt was measured. Then, using three methods of random forest, decision tree and linear regression, modeling was done to estimate CEC value. With the help of 70% of the data, the model was designed and with the remaining 30%, the validation or testing of the model was done. The correlation results showed that the CEC has a positive and significant correlation with the amount of soil clay (0.73) and organic carbon (0.46) and a negative and significant correlation with the amount of sand (-0.56). The results of the comparison of models showed that the random forest is the most suitable model for CEC with an explanation coefficient of 0.60 and RMSE of 2.658 respectively.

Keywords: Cation exchange capacity, random forest algorithm, decision tree