

## برآورد توان تولیدی رویشگاه راش شرقی در جنگل‌های هیرکانی با استفاده از الگوریتم درخت طبقه‌بندی و رگرسیون

احسان فکور<sup>۱</sup>، سید جلیل علوی<sup>۲\*</sup>، مسعود طبری کوچکسرایبی<sup>۳</sup>، کورش احمدی<sup>۴</sup>

۱. دانشجوی کارشناسی ارشد جنگلداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس

۲. استادیار گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس

۳. استاد گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس

۴. دانشجوی دکتری جنگلداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۹۴/۰۴/۰۵، تاریخ پذیرش: ۱۳۹۴/۱۱/۲۰

### چکیده

در مطالعه حاضر از شاخص فرم رویشگاه که مطمئن‌ترین معیار برای ارزیابی کیفیت رویشگاه در توده‌های ناهمسال و آمیخته است، برای ارزیابی توان تولیدی رویشگاه راش شرقی استفاده شد. به این منظور در جنگل پژوهشی دانشگاه تربیت مدرس، در تیپ‌هایی که در آنها گونه راش غالب بود، به روش منظم - تصادفی ۱۰۵ قطعه نمونه دایره‌ای به مساحت ۱۰۰۰ متر مربع پیاده و در هر یک از آنها ارتفاع و قطر تمام درختان گونه راش علاوه بر ارتفاع از سطح دریا، درصد شیب و آزیموت اندازه‌گیری و ثبت شد. همچنین از عمق ۱۰-۰ سانتی‌متری، نمونه‌برداری خاک انجام گرفت و متغیرهای فیزیکی و شیمیایی خاک اندازه‌گیری شد. ارزیابی توان تولید رویشگاه راش با استفاده از تکنیک درخت طبقه‌بندی و رگرسیون نشان داد که متغیرهای فسفر، شاخص تابش خورشیدی، درصد رس و وزن مخصوص ظاهری به ترتیب اهمیت نسبی، متغیرهای مؤثر بر فرم رویشگاه هستند و ۶۲ درصد تغییرات در توان تولیدی را می‌توان با استفاده از این متغیرها تبیین کرد. با استفاده از مدل خطی تعمیم‌یافته و معیارهای ضریب تبیین تعدیل‌یافته، مجذور مربعات میانگین خطا، معیارهای اطلاعاتی آکائیک (AIC) و بیزی (BIC) عملکرد تکنیک درخت طبقه‌بندی و رگرسیون نیز ارزیابی شد. نتایج نشان داد هرچند تکنیک درخت طبقه‌بندی و رگرسیون و مدل خطی تعمیم‌یافته درصد تغییرات یکسانی را توجیه می‌کند، الگوریتم درخت تصمیم از لحاظ معیارهای اطلاعاتی AIC و BIC نسبت به مدل رگرسیونی عملکرد بهتری دارد. از طرف دیگر تفسیر این تکنیک‌ها آسان‌تر است.

**واژگان کلیدی:** الگوریتم درخت طبقه‌بندی و رگرسیون، توان تولید رویشگاه، جنگل‌های هیرکانی، مدل خطی تعمیم‌یافته.

### مقدمه

مختلف جنگل همانند تولید چوب، ترسیب کربن، حفظ خاک، کنترل سیلاب، حفظ تنوع زیستی و تأمین آب نشان می‌دهد [۱]. از نظر بوم‌شناختی، توان تولید، پتانسیل یک محل برای تولید خالص نخستین را منعکس می‌کند، در صورتی که از دیدگاه جنگلداری عملی، پتانسیل آن را برای تولید چوب برای یک گونه مشخص بازگو می‌کند. بنابراین با برآورد قابلیت رویشگاه، اراضی جنگلی

ارزیابی توان تولید رویشگاه‌های جنگلی مقوله بسیار مهمی در مطالعات جنگلداری در دنیا است که در کشور ما کمتر به آن توجه شده و آن عبارت است از ارائه شاخصی که پتانسیل رویشگاه را در ایفای نقش‌های

\* نویسنده مسئول، شماره تماس: ۰۹۱۱۱۵۸۰۰۹۷

در خصوص ارزیابی توان تولید (با استفاده از شاخص توان تولید) در زمینه متغیرهای محیطی مطالعات بسیار اندکی صورت گرفته است و تنها می‌توان به منابع موجود زیر اشاره کرد. هررا و همکاران (۱۹۹۹) توان تولیدی رویشگاه *Vochycia ferruginea* را در جنگل‌های بارانی بررسی کردند [۵]. فرناندز و همکاران (۲۰۰۴) از مدل رگرسیون چندگانه برای ارتباط فرم رویشگاه با متغیرهای خاکی استفاده کردند [۶]. براوو اویدو و مونترو (۲۰۰۵) ارتباط شاخص رویشگاه و متغیرهای خاکی را برای گونه کاج سنگی (*Pinus pinea*) با استفاده از تحلیل رگرسیون بررسی کردند [۷]. مویسن و همکاران (۲۰۰۶) به برآورد حضور و سطح مقطع گونه‌های مختلف با استفاده از تکنیک‌های مدل‌سازی درخت رگرسیون تقویت‌شده<sup>۳</sup> (BRT) و مدل جمعی تعمیم‌یافته<sup>۴</sup> (GAM) و مقایسه این تکنیک‌ها با تکنیک‌های مرسوم درختان تصمیم در جنگل‌های کوهستانی یوتا در آمریکا پرداختند و به این نتیجه رسیدند که BRT بهترین برآورد را دارد [۸]. آرتسن و همکاران (۲۰۱۰) تکنیک‌های مختلف مدل‌سازی (BRT, CART, MLR) را برای پیش‌بینی شاخص رویشگاه در دو منطقه جنگل‌های کوهستانی مدیترانه‌ای در جنوب ترکیه با استفاده از متغیرهای توپوگرافی و اراضی جلگه‌ای در منطقه فلاندر در شمال بلژیک با استفاده از متغیرهای مربوط به خاک و هوموس بررسی و مقایسه کردند [۹].

در زمینه به‌کارگیری تکنیک درخت رگرسیون و طبقه‌بندی در جنگل نیز می‌توان به منابع زیر اشاره کرد. بیکر و همکاران (۱۹۹۳) به برآورد خطر مرگ‌ومیر *Pinus taeda* و *eliottii* در اثر *Heterobasidion annosum* با استفاده از تکنیک مدل‌سازی CART پرداختند [۱۰]. آیورسن و پراساد (۱۹۹۹) روابط بین گونه‌ها و محیط را در شرایط فعلی و تحت سناریوی تغییر اقلیم در آینده با

حاصلخیز باکیفیت را می‌توان به صورت متمرکز و اراضی جنگلی کم‌بازده و فقیر با کیفیت رویشگاهی کم را به صورت گسترده مدیریت کرد [۲].

تجزیه و تحلیل داده‌های پیچیده اکولوژیکی نیازمند روش‌های تحلیلی قوی و انعطاف‌پذیری است که بتواند روابط غیرخطی، تأثیرات متقابل و داده‌های ازدست‌رفته را کنترل کند. علاوه بر این، درک و ارائه نتایج از طریق این روش‌ها باید ساده و به راحتی تفسیرپذیر باشد. درخت تصمیم ابزاری قدرتمند و در عین حال رایج در داده‌کاوی است که هم برای طبقه‌بندی و هم برای پیش‌بینی استفاده می‌شود. جذابیت روش‌های درخت‌منا<sup>۱</sup> بیش از هر چیز به این دلیل است که به راحتی برای همگان قابل فهم است. از آنجا که درخت تصمیم، بررسی داده و مدل‌سازی را با هم ترکیب می‌کند، گام اولیه قدرتمندی در فرایند مدل‌سازی به‌شمار می‌رود [۳].

درخت تصمیم، ساختاری است که برای تقسیم مجموعه بزرگی از داده‌های جمع‌آوری‌شده به مجموعه‌های کوچک‌تر زنجیروار داده‌ها، براساس یک رشته قوانین ساده تصمیم استفاده می‌شود. در هر دسته‌بندی متوالی، اعضای مجموعه‌های حاصل بیش از پیش به یکدیگر شباهت پیدا می‌کنند [۴]. درخت طبقه‌بندی و رگرسیون<sup>۲</sup> (CART) برای تجزیه و تحلیل داده‌های پیچیده اکولوژیکی ایده‌آل و مناسب‌اند. این تکنیک یکی از روش‌های رگرسیون ناپارامتری است که درخت تصمیم را براساس الگوریتم دسته‌بندی دودویی بازگشتی رشد می‌دهد. درخت به صورت گرافیکی با گره ریشه (حاوی داده‌های تقسیم‌نشده) در بالا و گره‌های شاخه (داده‌های تقسیم‌شده) و گره‌های برگ (تقسیم‌های نهایی) در پایین، نشان داده می‌شود که این موضوع به درک و تفسیر روابط کمک می‌کند.

3. Boosted Regression Trees  
4. Generalized Additive Model

1. Tree-based methods  
2. Classification and Regression Tree

گرفته است. منطقه مورد مطالعه در سری ۳ حوضه ۴۶ کجور واقع است. به منظور جمع‌آوری اطلاعات مورد نیاز از ۱۰۵ قطعه نمونه دایره‌ای با مساحت ۱۰۰۰ متر مربع در توده‌هایی که در آنها گونه راش غالب بوده است، استفاده شد. پس از پیاده کردن قطعات نمونه در جنگل، قطر برابر سینه و همچنین ارتفاع کامل تمام درختان راش با قطر بیشتر از ۷/۵ سانتی‌متر اندازه‌گیری شد. سپس ویژگی‌های عمومی رویشگاه مثل ارتفاع از سطح دریا، درصد شیب و آزیموت جهت یادداشت شد. در مورد جهت جغرافیایی که دارای توزیع دایره‌ای است، نمی‌توان از داده‌های خام آزیموت در تجزیه و تحلیل استفاده کرد. به همین دلیل روش‌های مختلفی برای تبدیل آزیموت به یک متغیر کمی ابداع شده است [۱۸]. جهت جغرافیایی با استفاده از رابطه ۱ به شاخص تابش خورشیدی<sup>۱</sup> تبدیل شد که در آن  $\theta$  مقدار آزیموت جهت برحسب درجه است. مقدار شاخص تابش خورشیدی بین صفر و یک است و جهت شمال شرقی دارای مقدار صفر (خنک‌ترین دامنه) و جهت جنوب غربی دارای مقدار یک (گرم‌ترین دامنه) است [۱].

$$TRASP = [1 - \cos((\pi/180)(\theta - 30))] / 2 \quad (1)$$

در مرحله بعد در هر قطعه نمونه، نمونه‌برداری خاک از عمق ۰ تا ۱۰ سانتی‌متر صورت گرفت. نمونه‌های خاک پس از انتقال به آزمایشگاه در هوای آزاد خشک شده و از الک ۲ میلی‌متری عبور داده شد و پارامترهای بافت خاک به روش هیدرومتری، وزن مخصوص ظاهری به روش کلوخه و پارافین، pH خاک با دستگاه pH متر الکتریکی، نیتروژن به روش کج‌لدال، کربن آلی به روش اکسیداسیون دی‌کرومات، پتاسیم به وسیله دستگاه طیف‌سنج اتمی و فسفر قابل جذب به روش اولسن در آزمایشگاه آنالیز شد.

#### شاخص فرم رویشگاه

در این پژوهش از معیار فرم رویشگاه به‌عنوان معیاری از

استفاده از CART بررسی و متغیرهای مختلف را براساس اهمیت طبقه‌بندی کردند [۱۱]. هریسون و همکاران (۲۰۰۶) با استفاده از الگوریتم CART متغیرهای مختلف مؤثر بر پراکنش گونه‌های گیاهی را بررسی کردند [۱۲]. کالی و کونرت (۲۰۰۶) مدل درخت طبقه‌بندی و رگرسیون را جایگزینی برای ارزیابی خطر گونه‌های مهاجم در استرالیا معرفی کردند [۱۳]. اسپاداوپیا و همکاران (۲۰۰۸) از درخت طبقه‌بندی و رگرسیون برای انتخاب بهترین متغیرها برای طبقه‌بندی براساس شاخص سطح برگ استفاده کردند [۱۴]. یوسف و همکاران (۲۰۱۵) به منظور تهیه نقشه حساسیت به زمین‌لغزش از مدل‌های مختلف داده‌کاوی استفاده کردند و نتایج نشان داد که روش‌های جنگل‌های تصادفی، مدل جمعی تعمیم‌یافته، درخت رگرسیون تقویت‌شده و مدل خطی تعمیم‌یافته دقت مناسبی برای پیش‌بینی نقاط حساس به زمین‌لغزش دارند [۱۵]. نقیسی و همکاران (۲۰۱۵) از مدل‌های مختلف برای پتانسیل‌یابی آب زیرزمینی استفاده کردند که از بین این مدل‌ها روش‌های درخت مبنا مانند جنگل‌های تصادفی و درخت طبقه‌بندی و رگرسیون نتایج بهتری داشتند [۱۶].

توان تولید یک مؤلفه کلیدی از خدمات بوم‌سازگان جنگل بوده و معیار مهمی برای برنامه‌ریزی و مدیریت پایدار جنگل‌هاست، ولی در ایران در این زمینه کمتر تحقیق شده است، از این رو تحقیق حاضر قصد دارد توان تولیدی رویشگاه راش شرقی را در زمینه مشخصه‌های فیزیکی و شیمیایی خاکی و فیزیوگرافی با استفاده از الگوریتم داده‌کاوی درخت طبقه‌بندی و رگرسیون، در جنگل تحقیقاتی دانشگاه تربیت مدرس ارزیابی کند.

#### مواد و روش‌ها

##### منطقه تحقیق

مطالعه حاضر در جنگل آموزشی و پژوهشی صلاح‌الدین کلا متعلق به دانشکده منابع طبیعی دانشگاه تربیت مدرس انجام

تقسیم می‌شود و آن گره‌ها نیز پس از آن تقسیم می‌شوند، و به همین ترتیب این روند ادامه پیدا می‌کند [۱۸]. این تقسیم‌بندی تا زمانی ادامه می‌یابد که گروه‌ها همگن شوند یا تعداد مشاهدات در درون گره‌ها کمتر از مقدار حداقل تعریف شده توسط کاربر باشد. هر گره از طریق تعداد مشاهدات، مقدار میانگین متغیر پاسخ آن مشاهدات و میزان متغیر پیشگو تقسیم‌کننده آن مشخص می‌شود [۱۹].

یک موضوع مهم در درخت رگرسیون و طبقه‌بندی، تصمیم در مورد زمان توقف تقسیم‌هاست، چنانچه تقسیم شدن‌ها متوقف نشود، الگوریتم درخت در نهایت همه اطلاعات را تقسیم می‌کند و به بیش‌برازش یا فرایادگیری<sup>۱</sup> منجر می‌شود [۱۹]. به‌عنوان یک قاعده می‌توان گفت درخت پیچیده‌تر، دقت بیشتری دارد، اما درخت‌هایی که بیش از حد پیچیده باشند، ناپایدارند، به‌عبارت دیگر، با خطاهای بیشتری برای مشاهدات جدید همراه خواهند بود و نمی‌توان برای پیش‌بینی موارد دیگر از آنها استفاده کرد. از طرف دیگر، روشن است که یک درخت خیلی ساده نیز امکان رسیدن به پیش‌بینی خوب را فراهم نمی‌کند، از این‌رو باید درخت طبقه‌بندی و رگرسیون هرس شود. اندازه بهینه درخت را می‌توان با استفاده از معیار پارامتر پیچیدگی و تکنیک اعتبارسنجی متقابل k-fold تعیین کرد [۲۰].

از مدل خطی تعمیم‌یافته با توزیع گوسی و تابع پیوند همانی و با استفاده از معیارهای مجذور میانگین مربعات خطا (RMSE)، ضریب تبیین تعدیل‌یافته ( $R^2_{adj}$ )، آماره آکائیک<sup>۲</sup> (AIC) و معیار اطلاعاتی بی‌زی<sup>۳</sup> (BIC) برای قضاوت عملکرد درختان طبقه‌بندی و رگرسیون استفاده شد [۱]. به‌منظور اعتبارسنجی تکنیک درخت رگرسیون و مدل خطی تعمیم‌یافته نیز از روش اعتبارسنجی متقابل-10 fold در نرم‌افزار آماری R بهره گرفته شد.

به‌منظور اجرای تکنیک CART از نرم‌افزار آماری R

توان تولیدی رویشگاه استفاده شد [۲]. در توده‌های ناهمسال و آمیخته، ارتفاع در رابطه با سن (شاخص رویشگاه) نمی‌تواند شاخص خوبی برای بیان کیفیت رویشگاه باشد. در این جنگل‌ها از شاخص فرم رویشگاه استفاده می‌شود. فرم رویشگاه شاخصی از توان تولیدی رویشگاه برای جنگل‌های ناهمسال و آمیخته است و علاوه بر ارتفاع کل از قطر برابرسینه هم برای بررسی استفاده می‌کند.

از آنجا که برآورد فرم رویشگاه مبتنی بر مدلسازی رابطه قطر و ارتفاع است، باید مدلی که این رابطه را به بهترین وجه توصیف کند، انتخاب کرد. براساس مطالعات صورت‌گرفته مدل Chapman-Richards بهترین مدل در توصیف رابطه قطر و ارتفاع گونه راش در منطقه تحقیق است [۱۷]. براساس این مدل و روش منحنی راهنما یا معادله اختلاف، معادله یا منحنی‌های شاخص توان تولید برای گونه راش تهیه شده است. در مرحله بعد براساس میانگین ارتفاع و قطر برابر سینه گونه راش در هر قطعه نمونه، شاخص توان تولید محاسبه شد [۲].

### درختان طبقه‌بندی و رگرسیون

در پژوهش حاضر، برای ارزیابی توان تولیدی رویشگاه راش شرقی به کمک معیار فرم رویشگاه نسبت به متغیرهای محیطی از روش درختان طبقه‌بندی و رگرسیون استفاده شد. ساختار سلسله‌مراتبی درخت رگرسیون و طبقه‌بندی بدان معناست که پاسخ به یک متغیر به مقادیر در گره‌های بالاتر نیز بستگی دارد، بنابراین اثر متقابل بین متغیرها به‌طور خودکار مدلسازی می‌شود [۱].

اولین مرحله در مدلسازی CART، تصمیم در مورد انتخاب بهترین متغیر تقسیم‌کننده است که بتواند متغیر هدف را به بهترین نحو تقسیم کند. پس از انتخاب بهترین متغیر تقسیم‌کننده، داده‌ها به دو گروه تقسیم می‌شوند (تقسیم دودویی). جنبه تقسیم‌بندی بازگشتی به این واقعیت اشاره دارد که هر گره خود، به دو گره فرزند

1. Over-fitting  
2. Akaike Information Criteria  
3. Bayesian Information Criteria

تولید در هر گره آورده شده است که در آن متغیر درصد رس به‌عنوان متغیر اول برای طبقه‌بندی و فسفر و وزن مخصوص ظاهری به‌عنوان متغیر دوم انتخاب شد. همان‌طور که مشاهده می‌شود، بیشترین میزان توان تولید در رویشگاه‌هایی با درصد رس کمتر از ۳۸ درصد و وزن مخصوص ظاهری کمتر از ۱/۶ گرم بر سانتی‌متر مکعب و تابش خورشیدی بیشتر از ۰/۲ رخ می‌دهد. میانگین توان تولید در این رویشگاه‌ها بیشتر از ۳۶ متر است. از طرف دیگر، کمترین میزان توان تولید در رویشگاه‌هایی با درصد رس بیشتر از ۳۸ درصد و فسفر بیشتر از ۱۶ میلی‌گرم بر کیلوگرم مشاهده می‌شود.

در جدول ۲ اهمیت نسبی متغیرهایی که در تشکیل درخت دخالت داشته‌اند، ارائه شده است. با مراجعه به این جدول مشاهده می‌شود مهم‌ترین متغیر تأثیرگذار بر توان تولید رویشگاه راش متغیر فسفر و کم‌اهمیت‌ترین آن درصد سیلت است.

3.1.2 و بسته rpart استفاده شد. ابتدا درخت رگرسیون هرس نشده ترسیم و سپس برای هرس کردن درخت مقدار بهینه پارامتر پیچیدگی محاسبه شد. در نهایت مدل CART براساس مقدار بهینه پارامتر پیچیدگی هرس و ترسیم شد.

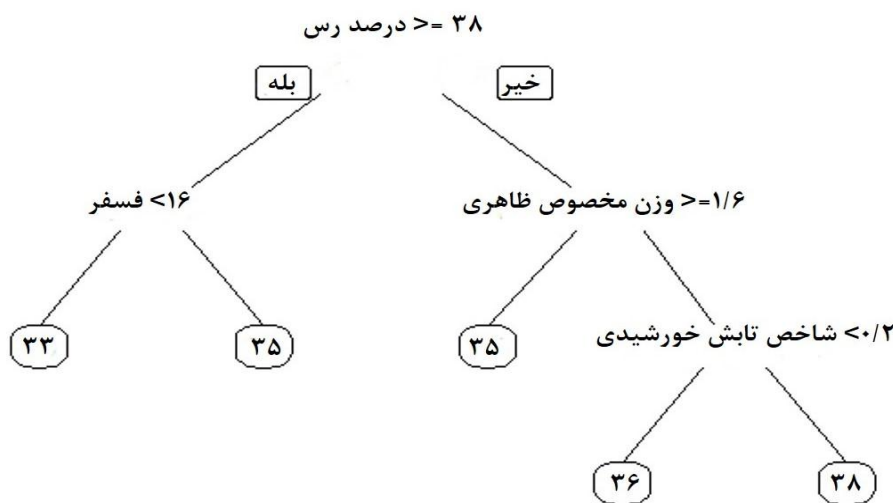
## نتایج و بحث

در جدول ۱ آماره‌های حداقل، حداکثر، میانگین، انحراف استاندارد و خطای استاندارد برای فرم رویشگاه و عوامل خاکی و فیزیوگرافی مورد استفاده برای پیش‌بینی توان تولیدی رویشگاه راش، آورده شده است.

هرس درخت کامل با استفاده از پارامتر پیچیدگی و اصل میانگین‌گیری نشان داد که مقدار میانگین پارامتر پیچیدگی با ۱۰۰ بار اجرای مدل ۰/۰۴۴ برآورد شد و براساس این مقدار، درخت رگرسیون هرس شد. در شکل ۱ درخت رگرسیون هرس شده به‌همراه مقدار میانگین توان

جدول ۱. مقادیر حداقل، حداکثر، میانگین، اشتباه معیار و انحراف استاندارد برای متغیرهای پاسخ و تبیینی

متغیرهای پاسخ و تبیینی	حداقل	حداکثر	میانگین	انحراف استاندارد	خطای استاندارد
فرم رویشگاه	۲۹/۲۵	۳۹/۶۶	۳۴/۸۶	۲/۳۴	۰/۲۰۷
میانگین قطر (سانتی‌متر)	۱۵/۱۴	۶۸/۵	۳۸/۸۳	۱۱/۳	۱/۰۰
میانگین ارتفاع (متر)	۱۹/۴۹	۳۷/۱۳	۲۷/۶۵	۴/۲۷	۰/۳۷
تعداد در پلات	۴	۵۱	۱۷	۸/۲۴	۰/۷۳
ارتفاع از سطح دریا	۱۰۶۷	۱۴۴۵	۱۲۲۸	۸۰/۶۳	۷/۱۵
شاخص تابش خورشیدی	۰	۱	۰/۳۰۳	۰/۲۸۲	۰/۰۲۵
درصد شن	۴	۶۲	۲۷/۳۳	۱۲/۲۴	۱/۰۸۶
درصد سیلت	۱۶	۵۶	۳۶/۸۱	۷/۶۷	۰/۶۸۱
درصد رس	۱	۶۴	۳۵/۸۵	۱۲/۰۳	۱/۰۶۸
وزن مخصوص ظاهری (g/cm <sup>3</sup> )	۱/۰۱۲	۲/۰۴۹	۱/۵۲	۰/۳۴۸	۰/۰۲۲
درصد ماده آلی	۱/۹۶	۱۳/۳۷	۶/۰۱۲	۳/۱۲	۰/۳۷۷
درصد کربن	۱/۱۳	۷/۷۶	۳/۴۸	۱/۸۱	۰/۱۶۱
درصد رطوبت اشباع	۴۳/۶۴	۵۵/۴۲	۴۸/۷۴	۲/۶۱	۰/۲۳۲
پتاسیم (mg/kg)	۶	۲۲۴	۹۹/۲۰	۶۳/۴۶	۵/۶۳
فسفر (mg/kg)	۴	۳۷/۸۴	۱۴/۱۳	۸/۷۹	۰/۷۸
شیب (درصد)	۳/۳	۷۲/۶	۲۶/۱۹	۱۲/۳۷	۱/۰۹
درصد نیتروژن	۰/۱۳	۰/۶۹	۰/۳۲	۰/۱۱	۰/۰۱
نسبت کربن به نیتروژن	۴/۴۶	۲۰/۱۸	۱۰/۲۸	۳/۲۵	۰/۲۸
اسیدیته	۵/۱	۷/۵۳	۵/۹۸	۰/۵۴	۰/۰۴۸



شکل ۱. درخت رگرسیون هرس شده

جدول ۲. اهمیت نسبی متغیرهای پیشگو در تکنیک درخت طبقه‌بندی و رگرسیون

متغیر پیشگو	اهمیت نسبی	نسبت کربن به نیتروژن	درصد رس	شاخص تابش خورشیدی	وزن مخصوص ظاهری	اسیدیته	ارتفاع از سطح دریا	پتاسیم	درصد سیلت
اهمیت نسبی	۰/۷۴	۰/۵۶	۰/۵۴	۰/۴۹	۰/۴۶	۰/۳۶	۰/۳۰	۰/۲۱	۰/۱۹

جدول ۳. معیارهای ارزیابی مدل برای تکنیک درخت رگرسیون و مدل خطی تعمیم‌یافته

مدلسازی				اعتبارسنجی				مدل
BIC	AIC	RMSE	R <sup>2</sup> adj	BIC	AIC	RMSE	R <sup>2</sup> adj	
۷۸/۲۲	۳۸/۴۱	۱/۰۸	۶۸/۰۱	۹۳/۰۹	۵۳/۲۸	۱/۲۵	۴۸/۸۶	مدل خطی تعمیم‌یافته
۲۵/۷۴	۲۵/۱۳	۱/۱۸	۶۲/۲۴	۶۱/۹۲	۵۱/۳۰	۱/۵۱	۴۳/۹۳	تکنیک درخت رگرسیون

پژوهش‌های [۱، ۸] که نشان دادند مدل رگرسیون خطی عملکرد بهتری نسبت به درخت تصمیم دارد مغایر است، هرچند در بعضی پژوهش‌ها به توانایی بهتر مدل‌های ناپارامتری در پیش‌بینی متغیرهای پاسخ در داده‌های اکولوژیک و مشخصات کمی جنگل اشاره شده است.

استفاده از دامنه گسترده‌ای از انواع متغیرهای پاسخ از جمله متغیرهای کمی و طبقه‌ای، تغییرناپذیری به تبدیل متغیرهای تبیینی، سهولت ساخت درختان، سهولت تفسیر و غیرحساس بودن به نقاط دورافتاده، از جمله مزیت‌های درخت تصمیم است [۲۰]. در مطالعه حاضر برای بررسی دقیق‌تر عملکرد تکنیک درخت رگرسیون، تأثیرات متقابل

مقایسه مدل‌ها با استفاده از رایج‌ترین معیارهای ارزیابی مدل همانند معیارهای  $R^2_{adj}$ ، RMSE، AIC و BIC و اعتبارسنجی مدل‌ها با استفاده از تکنیک 10-fold نشان داد که از نظر معیارهای  $R^2_{adj}$  و RMSE اختلاف اندکی بین دو روش وجود دارد، اما با به‌کارگیری معیارهای اطلاعاتی AIC و BIC که تعادلی میان دقت مدل و پیچیدگی آن برقرار می‌کند، حاکی از آن است که اختلاف دو مدل رگرسیون خطی و درخت تصمیم چشمگیر است. از این رو می‌توان گفت که تکنیک درخت تصمیم با توجه به جمع معیارهای ارزیابی، کارایی بهتری نسبت به مدل خطی تعمیم‌یافته دارد (جدول ۳). نتایج این تحقیق با نتایج

احمدی و همکاران [۲] که بیان کردند درصد سیلت کم‌اهمیت‌ترین متغیر و نسبت کربن به ازت یکی از مهم‌ترین متغیرها در ارزیابی توان تولیدی رویشگاه راش با استفاده از مدل جمعی تعمیم‌یافته است، تطابق دارد.

### نتیجه‌گیری

تحلیل داده‌های پیچیده بوم‌شناختی نیازمند روش‌های تحلیلی انعطاف‌پذیر و قوی است که بتواند روابط غیرخطی و تأثیرات متقابل را کنترل کند. علاوه بر این، درک و ارائه نتایج از طریق این روش‌ها باید ساده و به راحتی تفسیرپذیر باشد. شایان ذکر است که هرچند ساخت درختان رگرسیون و طبقه‌بندی آسان و تفسیر آنها نیز به راحتی امکان‌پذیر است، یک ضعف اساسی آن این است که مدل برازش‌یافته واریانس بالایی دارد. تغییر اندک در داده‌ها (با جمع‌آوری یک نمونه متفاوت از داده‌ها و برازش مدل) اغلب به تغییرات زیاد در شکل درخت برازش‌یافته منجر می‌شود. این امر تفسیر درخت رگرسیون و طبقه‌بندی را با مشکل مواجه می‌کند. برای غلبه بر این مشکل، راه‌حل‌های متعددی ارائه شده که تمامی این روش‌ها مستلزم برازش درختان متفاوت زیاد به داده‌ها و میانگین‌گیری پیش‌بینی‌ها از درختان است. این رویکردها که به روش ترکیبی معروف‌اند، در علوم جنگل و اکولوژی بسیار کمتر از سایر تکنیک‌ها استفاده شده است. از این رو پیشنهاد می‌شود در مطالعات آینده از این تکنیک‌ها و به خصوص جنگل تصادفی و درخت رگرسیون تقویت‌شده در ارزیابی توان تولیدی رویشگاه‌های جنگلی استفاده شود.

دو طرفه از متغیرهای پیشگو به مدل خطی تعمیم‌یافته اضافه شد. نتایج نشان داد که این مدل ۶۸ درصد از تغییرات در توان تولید را توجیه کرده است، اما برازش مدل خطی بدون در نظر گرفتن اثر متقابل، ضریب تبیین ۴۷ درصدی را نشان می‌دهد. تکنیک‌های معمول مدلسازی آماری اغلب در یافتن الگوهای معنادار اکولوژیک از چنین داده‌هایی ناتوان‌اند، اما تکنیک‌های درخت تصمیم ابزار قدرتمندی بدین منظورند.

رویکرد درخت رگرسیون قواعدی را ایجاد می‌کند که در نهایت می‌تواند به طبقه‌بندی توان تولیدی رویشگاه منجر شود. الگوریتمی که این قواعد را می‌سازد، مقدار دقیق متغیر پیشگو را در هر شاخه از درخت رگرسیونی شناسایی می‌کند. به کارگیری درخت رگرسیونی در مطالعه حاضر نشان داد که درصد رس، وزن مخصوص ظاهری، فسفر و شاخص تابش خورشیدی در تعیین توان تولیدی رویشگاه راش تأثیرگذار و مهم بوده‌اند. مسئله مهمی که باید به آن توجه داشت این است که هرچند در هر مرحله از شکافتن داده، تنها یکی از متغیرها بهترین تفکیک را در داده ایجاد می‌کند، ممکن است متغیرهای دیگری نیز وجود داشته باشند که داده‌ها را به خوبی تقسیم کنند. در بعضی موارد بررسی سایر متغیرها، تفسیر نتایج را آسان‌تر می‌کند. همان‌طور که اشاره شد، متغیرهای فسفر، وزن مخصوص ظاهری، درصد رس، شاخص تابش خورشیدی، وزن مخصوص ظاهری، اسیدیته، ارتفاع از سطح دریا، پتاسیم و درصد سیلت به ترتیب مهم‌ترین و کم‌اهمیت‌ترین متغیرها در تعیین توان تولیدی رویشگاه راش شرقی بوده‌اند. نتایج پژوهش حاضر با نتایج تحقیق

### References

- [1]. Aertsen, W., Kint, V., Van Orshoven, J., and Muys, B. (2011). Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environmental Modelling & Software*, 26(7): 929-937.
- [2]. Ahmadi, k., Alavi, S.J., and Tabari Kouchaksaraei, M. (2015). Evaluation of oriental beech (*fagus orientalis lipsky*.) site productivity using generalized additive model (case study: Tarbiat Modares University forest research station). *Journal of Iranian Forest*, 7(1): 17-32.

- [3]. Crawford, S.L. (1989). Extensions to the CART algorithm. *International Journal of Man-Machine Studies*, 31(2): 197-217.
- [4]. Simpson, G.L., and Birks, H.J.B. (2012). *Statistical learning in palaeolimnology*. In *Tracking environmental change using lake sediments*, Springer, Netherlands.
- [5]. Herrera, B., Campos, J.J., Finegan, B., and Alvarado, A. (1999). Factors affecting site productivity of a Costa Rican secondary rain forest in relation to *Vochysia ferruginea*, a commercially valuable canopy tree species. *Forest Ecology and Management*, 118(1-3): 73-81.
- [6]. Herrera-Fernández, B., Campos, J.J., and Kleinn, C. (2004). Site productivity estimation using height-diameter relationships in Costa Rican secondary forests. *Investigación Agraria. Sistemasy Recursos Forestales*, 13(2): 295-303.
- [7]. Bravo-Oviedo, A., and Montero, G. (2005). Site index in relation to edaphic variables in stone pine (*Pinus pinea* L.) stands in south west Spain. *Annals of Forest Science*, 62(1): 61-72.
- [8]. Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., and Edwards Jr, T.C. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, 199(2): 176-187.
- [9]. Aertsens, W., Kint, V., Van Orshoven, J., Ozkan, K., and Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modeling*, 221(8): 1119-1130.
- [10]. Baker, F. A., Verbyla, D. L., Hodges Jr, C. S., and Ross, E. W. (1993). Classification and regression tree analysis for assessing hazard of pine mortality caused by *Heterobasidion annosum*. *Plant Disease*, 77(2): 136-139.
- [11]. Iverson, L.R., Prasad, A., and Schwartz, M.W. (1999). Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling*, 115(1): 77-93.
- [12]. Harrison, P.A., Berry, P.M., Butt, N., and New, M. (2006). Modelling climate change impacts on species' distributions at the European scale: implications for conservation policy. *Environmental Science & Policy*, 9(2): 116-128.
- [13]. Caley, P., and Kuhnert, P.M. (2006). Application and evaluation of classification trees for screening unwanted plants. *Austral Ecology*, 31(5), 647-655.
- [14]. Spadavecchia, L., Williams, M., Bell, R., Stoy, P.C., Huntley, B., and Van Wijk, M.T. (2008). Topographic controls on the leaf area index and plant functional type of a tundra ecosystem. *Journal of Ecology*, 96(6): 1238-1251.
- [15]. Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., and Al-Katheeri, M.M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, 13(5): 839-856.
- [16]. Naghibi, S.A., and Pourghasemi, H.R. (2015). A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resources Management*, 29(14): 5217-5236.
- [17]. Ahmadi, K., Alavi, S.J., Tabari Kouchaksaraei, M., and Aertsens, W. (2013). Non-linear height-diameter models for oriental beech (*Fagus orientalis* Lipsky) in the Hyrcanian forests, Iran. *Base*, 17(3): 431-440.
- [18]. Therneau, T.M., Atkinson, B., and Ripley, B. (2014). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-8. <http://CRAN.R-project.org/package=rpart>.
- [19]. Wilkinson, L. (1992). Tree structured data analysis: AID, CHAID and CART. In *Proceedings of Sawtooth Software Conference*.
- [20]. De'ath, G., and Fabricius, K.E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11): 3178-3192.



## Estimating the beech forest site productivity in Hyrcanian forest using classification and regression tree algorithm

**E. Fakour**; M.Sc. Student in Forestry, Faculty of Natural Resources, Tarbiat Modares University, Nour, I.R. Iran

**S. J. Alavi\***; Assist. Prof., Faculty of Natural Resources, Tarbiat Modares University, Nour, I.R. Iran

**M. Tabari Kouchaksaraei**; Prof., Faculty of Natural Resources, Tarbiat Modares University, Nour, I.R. Iran

**K. Ahmadi**; Ph.D. Student in Forestry, Faculty of Natural Resources, Tarbiat Modares University, Nour, I.R. Iran

(Received: 26 June 2015, Accepted: 09 February 2016)

### ABSTRACT

In this study, the site form index which is the most reliable criterion for evaluation of forest site productivity in uneven-aged and mixed stands was used. For this purpose, random-systematic sampling method was used to locate 105 0.1 ha circular sample plots in beech dominated forests in Tarbiat Modares University research forest. The height and diameter of *Fagus orientalis* Lipsky trees within each sample plot was recorded along with elevation, azimuth and slope of the ground. Also, at the center of plot, soil samples from first layer (0-10 cm) were taken for analyzing several soil variables. Evaluation of forest site productivity by using classification and regression tree algorithm showed that after pruning the full tree, phosphorus, TRASP, clay and bulk density are effective variables, in order of relative importance, on site form and 62% variations in productivity can be explained by these variables. Using generalized linear model and evaluation criteria such as AIC, RMSE,  $R^2$  and adjusted  $R^2$  the performance of CART model was assessed. The results showed though CART techniques and the generalized linear model justify the same variability in forest productivity but decision tree technique in terms of AIC and BIC criteria is better than the GLM and as well as this technique is easier to interpret.

**Keywords:** Classification and regression tree algorithm, Forest productivity, Generalized linear Model, Hyrcanian forests.

---

\* Corresponding Author, Email: j.alavi@modares.ac.ir, Tel: +98114425310