

ارزیابی کارایی مدل‌های داده‌کاوی در تعیین پراکنش گونه گیاهی *Anchusa strigosa* (مطالعه موردی: دزفول)

مریم اسدی*^۱، زهرا جعفری^۲

چکیده

بهره‌برداری از گیاهان صنعتی و دارویی بعنوان یکی از محصولات فرعی مراتع نقش موثری در اقتصاد مراتع دارد. حفظ و توسعه پراکنش گیاهان دارویی در سطح عرصه‌های مرتعی از جنبه‌های مختلف اقتصادی و اجتماعی و فنی مراتع، حایز اهمیت می‌باشد. در سطح مراتع گونه‌های گیاهی صنعتی و دارویی فراوانی موجود می‌باشد. گاوزبان خارک‌دار (*Anchusa strigosa*) از گونه‌های دارویی و بومی شهرستان دزفول می‌باشد. یکی از موارد حائز اهمیت در بهره‌برداری اصولی از این گیاه دارویی تعیین مناطق رویشگاهی آن می‌باشد. هدف از این تحقیق، بررسی رابطه ویژگی‌های بیوفیزیکی رویشگاه بالقوه این گونه و پراکنش آن با استفاده از تکنیک‌های مختلف داده‌کاوی است. برای ارزیابی کارایی مدل‌های داده‌کاوی، ابتدا ۳۰ نقطه حضور و ۳۰ نقطه عدم حضور گونه با استفاده از نقشه پوشش گیاهی و مطالعات میدانی ثبت شد. سپس پارامترهای ارتفاع، شیب، درجه حرارت، بارندگی، بافت و عمق خاک از نقشه‌های به دست آمده با استفاده از روش‌های زمین آمار بر مبنای داده‌های منطقه‌ای استخراج شد. تعیین ضریب تاثیر هر یک از پارامترهای ذکر شده با استفاده از روش ماشین بردار انجام شد. سپس اقدام به پیش‌بینی حضور گونه‌های مذکور براساس پارامترهای بیوفیزیکی با استفاده از روش‌های داده‌کاوی گردید. الگوریتم‌های داده‌کاوی مورد استفاده در این تحقیق شامل ازدحام ذرات، رگرسیون لجستیک، شبکه‌های عصبی مصنوعی و نزدیک‌ترین همسایه می‌باشند. نتایج نشان داد که الگوریتم ازدحام ذرات و شبکه عصبی مصنوعی با میزان دقت ۰/۹۵ و ضریب کاپای ۰/۹ دارای دقت بالاتری در تعیین مناطق حضور گونه می‌باشد. نتایج این تحقیق نشان داد با استفاده از روش‌های داده‌کاوی امکان پیش‌بینی حضور گونه مورد نظر، براساس پارامترهای بیوفیزیکی با دقت بالایی فراهم می‌باشد.

واژگان کلیدی: الگوریتم ازدحام ذرات، پیش‌بینی رویشگاه، رگرسیون لجستیک، شبکه‌های عصبی مصنوعی، نزدیک‌ترین همسایه.



مقاله پژوهشی

۱. دکتری آبخیزداری، اداره منابع طبیعی و آبخیزداری خوانسار، خوانسار، ایران.

m.asadi.96@ut.ac.ir

* نویسنده مسئول

۲. کارشناسی ارشد مرتعداری، گروه طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، ایران.

zahra.jafari16@yahoo.com

شناسه مقاله: ۲۲۰۶-۱۰۲۷
شماره صفحه پیاپی: ۲۳۴-۲۴۵

تاریخ دریافت: ۱۴۰۱/۰۳/۲۵
تاریخ پذیرش: ۱۴۰۱/۱۰/۲۵
انتشار آنلاین: ۱۴۰۱/۱۱/۰۲
زمان پذیرش: ۲۱۴ روز

استناددهی:

اسدی، م. و جعفری، ز. (۱۴۰۱). ارزیابی کارایی مدل‌های داده‌کاوی در تعیین پراکنش گونه گیاهی *Anchusa strigosa* (مطالعه موردی: دزفول). مدیریت اکوسیستم‌های طبیعی، (۳)، ۲۴-۳۵.

۱- مقدمه

ارزیابی و تحلیل اطلاعات مکانی گونه‌های گیاهی در علوم محیطی گامی مؤثر در برنامه‌ریزی و مدیریت یک اکوسیستم می‌باشد. شناسایی عوامل موثر بر پراکنش گونه‌های گیاهی و تشخیص ارتباط و برهم کنش‌های غیرخطی و پیچیده متغیرها یکی از موارد حائز اهمیت در این زمینه می‌باشد (Cutler et al., 2007). درک و شناسایی روابط پوشش گیاهی و عوامل محیطی موثر بر پراکنش گونه‌های گیاهی یکی از راهکارهای مدیریت عرصه‌های مرتعی و بهره‌برداری از گیاهان دارویی و صنعتی ارزشمند مراتع می‌باشد (پیری صحرا گرد و همکاران، ۱۳۹۴). امروزه با به کارگیری علوم جدید و روش‌های پیشرفته آماری می‌توان به کمی‌سازی روابط بین عوامل موثر بر پراکنش گونه‌های گیاه پرداخت. گونه‌های دارویی و صنعتی موجود در مراتع یکی از مهمترین منابع اقتصادی مراتع بوده که نقش موثری در اقتصاد مرتعداران و بهره‌برداران عرصه‌های مرتعی دارد. شناسایی رویشگاه گونه‌های صنعتی و دارویی با قدرت رویشی بالا و سازگار با شرایط محیطی یکی از موارد حائز اهمیت در افزایش راندمان تولیدات مرتعی می‌باشد. گونه‌های دارویی و صادراتی ارزش اقتصادی بالایی را به مراتع بخشیده‌اند. کشت گیاهان دارویی به ویژه در صورت امکان گیاهان دارویی بومی از بهترین روش‌های اصلاح مراتع می‌باشد که می‌توان از طریق آن هم مراتع را از نابودی نجات داد و هم موجبات اشتغال و درآمدزایی ساکنین منطقه را فراهم نمود.

در بوم‌شناسی کمی کردن رابطه بین گونه‌های گیاهی با عوامل محیطی بخشی از مدل‌های پیش‌بینی توزیع مکانی گونه‌هاست این مدل‌ها عموماً بر اساس این فرضیه استوار است که عوامل محیطی پراکنش گونه‌های گیاهی را کنترل می‌کنند (زارع چاهوکی، ۱۳۸۵). امروزه به کارگیری روش‌های آماری مناسب و سیستم اطلاعات جغرافیایی و مدل‌های پیش‌بینی توزیع جغرافیایی رویشگاه‌ها به سرعت در بوم‌شناسی توسعه یافته است (زارع چاهوکی، ۱۳۸۵). مدل‌سازی پیش‌بینی پراکنش گونه‌های گیاهی، به عنوان پیش‌بینی پراکنش بالقوه یک گونه گیاهی در سراسر چشم‌انداز، براساس ارتباط بین نقاط رخداد گونه گیاهی و متغیرهای محیطی مؤثر تعریف می‌شود (Franklin, 1995) و براساس این فرضیه است که عوامل محیطی پراکنش گونه گیاهی را کنترل می‌کنند (زارع چاهوکی، ۱۳۸۵). مدل‌های پیش‌بینی کننده رویشگاه، تناسب رویشگاه را برای استقرار گونه‌های گیاهی و جانوری مشخص می‌کنند (زارع چاهوکی، ۱۳۹۳) و به مدیران منابع طبیعی کمک می‌کند تا با اختصاص زمان و هزینه کمتر، به شناسایی عوامل تهدید کننده جمعیت‌ها بپردازند (Guisan and Zimmermann, 2000).

مدل‌های داده‌کاوی^۱ نیز ابزارهای جدیدی هستند که در چند دهه اخیر به منظور پیش‌بینی فرآیندهای مختلف در علوم منابع طبیعی و محیطی مورد استفاده قرار گرفته‌اند. برای انجام کارهایی مثل دسته‌بندی، پیش‌بینی، تخمین و خوشه‌بندی داده‌ها می‌توان از داده‌کاوی استفاده نمود. برای انجام این کارها الگوریتم‌های متعددی با ساختارهای متفاوت توسعه یافته‌اند که با توجه به پیشرفت رایانه‌ها و علوم داده‌کاوی همه روزه بر تعداد و کیفیت این الگوریتم‌ها افزوده می‌شود. که می‌توان به مواردی مانند الگوریتم خوشه‌بندی^۲، شبکه عصبی مصنوعی^۳، الگوریتم ژنتیک^۴، نزدیک‌ترین همسایه^۵ و درخت تصمیم^۶ اشاره نمود. در حقیقت داده‌کاوی فرآیند استخراج اطلاعات (قوانین و الگوهای) ضمنی، غیربدیهی، از قبل ناشناخته و بالقوه مفید از داده‌ها در پایگاه داده است (Frawley et al., 1992). تعاریف متنوعی از داده‌کاوی در مراجع مختلف ارائه شده است اما تعریف "استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده‌های بسیار بزرگ و پیچیده" در اکثر منابع دیده می‌شود (Asadi et al., 2021).

از جمله مطالعات انجام شده در خصوص شناسایی رویشگاه گونه‌های گیاهی می‌توان به مطالعه زارع چاهوکی و همکاران (۱۳۹۳) اشاره نمود که در مراتع شرق سمنان با استفاده از رگرسیون لجستیک^۷ به مدل‌سازی پراکنش گونه‌های گیاهی بر اساس عوامل خاک و توپوگرافی پرداختند. آنها به این نتیجه رسیدند که نقشه‌های پیش‌بینی رویشگاه گونه‌های گیاهی *Zygophyllum Halocnemum strobilaceum Astragalus spp.* و *Seidlitzia rosmarinus* بهتر از سایر گونه‌ها با نقشه پوشش گیاهی تطابق داشته ولی گونه *Artemisia sieberi* به دلیل داشتن دامنه وسیع بوم‌شناسی، نقشه پیش‌بینی با واقعیت زمینی تطابق مناسبی نداشت. بنابراین، مدل رگرسیون لجستیک توانایی ارائه نقشه‌ای را که با نقشه واقعی پوشش گیاهی این گونه گیاهی تطابق داشته باشد ندارد.

تقی‌زاده مهرجردی و همکاران (۱۳۹۳)، به نقشه‌برداری رقومی کلاس‌های خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه اردکان استان یزد پرداختند. نتایج نشان داد که مدل درخت تصمیم نسبت به سایر مدل‌ها دارای برتری نسبی می‌باشد به طوری که توانسته است دقت پیش‌بینی گروه‌های بزرگ خاک را نسبت به روش تحلیل تشخیصی (ضعیف‌ترین مدل) ۴۴٪ افزایش دهد. در کل نتایج تحقیق، دقت مدل‌های درخت تصمیم، شبکه عصبی مصنوعی، ترکیب شبکه عصبی مصنوعی - الگوریتم ژنتیک، رگرسیون لجستیک چند جمله‌ای^۸ و تحلیل تشخیصی^۹ را جهت پیش‌بینی گروه‌های بزرگ خاک با دقت کلی ۷۰٪، ۶۵٪، ۶۵٪، ۵۵٪ و ۴۷٪ به ترتیب تایید کرد.

1. Data Mining

2. Clustering algorithm

3. Artificial Neural Network

4. Genetic algorithm

5. Nearest Neighbor

6. Decision Tree

7. Logistic Regression (LR)

8. Multinomial Logistic Regression

9. Discriminant Analysis

پیری صحراگرد و همکاران (۱۳۹۴)، به ارزیابی کارایی مدل شبکه عصبی مصنوعی برای تهیه نقشه پیش‌بینی رویشگاه‌های گیاهی در مراتع استان قم پرداختند. بدین منظور، با روی هم گذاری نقشه‌های شیب، جهت و ارتفاع، واحدهای همگن تهیه شدند و نمونه‌برداری از پوشش گیاهی و خاک انجام شد. در آخر نقشه مربوط به متغیرهای محیطی با استفاده از سامانه اطلاعات جغرافیایی^۱ و زمین آمار^۲ تهیه شد. برای تهیه مدل شبکه عصبی از روش پرسپترون چندلایه^۳ بهره‌گیری شد. بر اساس نتایج، دقیق‌ترین مدل پیش‌بینی برای همه رویشگاه‌ها با استفاده از تابع انتقال تانژانت سیگموئید^۴ و قانون آموزش لوبنبرگ مارکوارت^۵ حاصل شد. نتایج نشان داد که نقشه‌های پیش‌بینی برای رویشگاه *Artemisia sieberi* دارای تطابق عالی؛ رویشگاه *Halocnemum strobilaceum* دارای تطابق خیلی خوب؛ رویشگاه *Tamarix passerinoides* دارای تطابق خوب؛ رویشگاه *Seidlitzia rosmarinus* دارای تطابق متوسط و رویشگاه *Artemisia sieberi* دارای تطابق ضعیف با نقشه‌های واقعیت زمینی است. این نتایج گویای آن است که شبکه پرسپترون چند لایه در مدل‌سازی و برآورد محدوده جغرافیایی پراکنش رویشگاه گونه‌های مورد مطالعه از دقت قابل قبولی برخوردار است و در صورتی که متغیرهای ورودی به شبکه به درستی انتخاب شوند می‌تواند شبیه‌سازی حضور و عدم حضور را با دقت بالایی انجام دهند.

فاطمی و همکاران (۱۳۹۶) پیش‌بینی پراکنش جغرافیایی گونه *Juniperus excelsa* با استفاده از داده‌های اقلیمی در شرایط حاضر و آینده در استان سمنان را مورد مطالعه قرار دادند. پراکنش گونه *Juniperus excelsa* در شرایط حال حاضر و آینده تحت اثر تغییر اقلیم (سال‌های ۲۰۵۰ و ۲۰۷۰) تحت دو سناریوی RCP2.6^۶ و RCP8.5 با سری داده مدل گردش عمومی HadGEM2_ES با استفاده از مدل توزیع گونه‌ای MaxEnt^۷ و ۲۲ متغیر اقلیمی و توپوگرافی، در ۱۶۰ سایت نمونه‌برداری استان سمنان، بررسی شده است. نتایج حاکی از آن است که پس از متغیر ارتفاع از سطح دریا، بارش سردترین فصل سال و درجه حرارت متوسط سالانه، مهمترین متغیرهای اثرگذار بر پراکنش گونه است. همچنین نتایج نشان داد که با ادامه روند کنونی گرمایش جهانی، سطح رویشگاه گونه نسبت به شرایط حاضر تحت سناریوی خوش بینانه RCP2.6 تا سال ۲۰۷۰، در طبقه رویشگاه متوسط ۸۲/۴ درصد افزایش، در طبقات رویشگاه خوب و بسیار خوب به ترتیب ۳۱/۳ و ۶۷/۳ درصد کاهش خواهد یافت و تحت سناریوی بدبینانه RCP8.5 در طبقات رویشگاه متوسط، خوب و بسیار خوب به ترتیب ۷۶/۵، ۹۸ و ۱۰۰ درصد با کاهش سطح مواجه خواهد شد. نتایج نشان داد که مدل‌هایی مانند MaxEnt با دقت قابل قبولی، کارایی لازم را برآورد پراکنش گونه دارند و متخصصان می‌توانند از آنها برای ارائه راهکارهای حفاظتی استفاده کنند.

پیری صحرا گرد (۱۳۹۶) به مدل‌سازی پیش‌بینی پراکنش رویشگاه گونه‌های گیاهی با روش رگرسیون لجستیک در مراتع غرب تفتان، شهرستان خاش پرداخت. در این مطالعه از رگرسیون لجستیک به عنوان مدل پیش‌بینی پراکنش رویشگاه گونه‌های گیاهی و به منظور استخراج نقشه پیش‌بینی پراکنش هر یک از رویشگاه‌ها استفاده شد. بر اساس مقادیر شاخص کاپا^۸، میزان تطابق نقشه‌های پیش‌بینی حاصل با نقشه واقعی برای رویشگاه *Artemisia aucheri* و *Amygdalus scoparia*، عالی؛ برای رویشگاه‌های *Heracleum persicum* و *Zygophyllum eurypterum* خوب و برای رویشگاه *Artemisia sieberi* ضعیف ارزیابی شد. نتایج نشان می‌دهد که روش رگرسیون لجستیک قادر است که برای رویشگاه گونه‌های *Artemisia aucheri* و *Amygdalus scoparia* که دارای شرایط رویشگاهی منحصر به فردی است مدل پیش‌بینی دقیقی فراهم آورد، اما برای رویشگاه *Artemisia sieberi*، به دلیل دامنه بوم‌شناختی گسترده، دقت مدل پیش‌بینی حاصل از این روش پایین بود.

دهقانی‌نژاد و همکاران (۱۳۹۷) به بررسی قابلیت پیش‌بینی مدل‌های آماری در ارزیابی توان تولید رویشگاه راش شرقی پرداخته و قابلیت پیش‌بینی مدل‌های خطی و جمعی تعمیم یافته برای ارتفاع غالب گونه راش به عنوان شاخصی عالی از کیفیت رویشگاه را مورد بررسی قرار دادند. نتایج نشان داد متغیر ارتفاع از سطح دریا به‌عنوان مهمترین متغیر اثرگذار بر ارتفاع غالب گونه راش شناسایی شد. سپس با استفاده از معیارهای ارزیابی حاصل از داده‌های مدل‌سازی مشاهده گردید مدل جمعی تعمیم‌یافته از نظر معیارهای ارزیابی، عملکرد بهتری نسبت به مدل خطی تعمیم یافته دارد.

Manel et al. (۱۹۹۹)، به مقایسه دو مدل شبکه‌های عصبی مصنوعی و رگرسیون لجستیک به منظور پیش‌بینی پراکنش مکانی رویشگاه بالقوه گونه *Rhycornis fuliginosus* پرداختند و با استفاده از سطح زیر منحنی نمودارهای مشخصه عملکرد^۹ ارزیابی صحت مدل^{۱۰} را انجام دادند. نتایج نشان داد که مدل رگرسیون لجستیک از دقت بیشتری نسبت به شبکه عصبی مصنوعی برخوردار است. با توجه به کارآمدی مدل‌های داده‌کاوی در پیش‌بینی‌ها با استفاده از برقراری ارتباط مناسب بین فرآیندهای متخلف در طبیعت، این مدل‌ها می‌توانند ابزار مناسبی در جهت پیش‌بینی مکان رویشی گونه‌ها براساس پارامترهای اقلیمی و فیزیوگرافی منطقه باشند.

در این مطالعه به بررسی کارایی مدل‌های داده‌کاوی در تعیین رویشگاه‌های گونه گیاهی دارویی گاوزبان خارک‌دار^{۱۱} در شهرستان دزفول پرداخته شده است بدین منظور در ابتدا با استفاده از سیستم اطلاعات جغرافیایی و عملیات‌های صحرایی داده‌های پوشش گیاهی و متغیرهای محیطی از جمله

1. Geographic Information System (GIS)

2. Geostatistics

3. Multi Layer Perceptron (MLP)

4. Tangent Sigmoid

5. Levenberg- Marquardt

6. Representative Concentration Pathways

7. The Maximum Entropy Model

8. Kappa Coefficient

9. Receiver Operating Characteristic (ROC)

10. Model validation

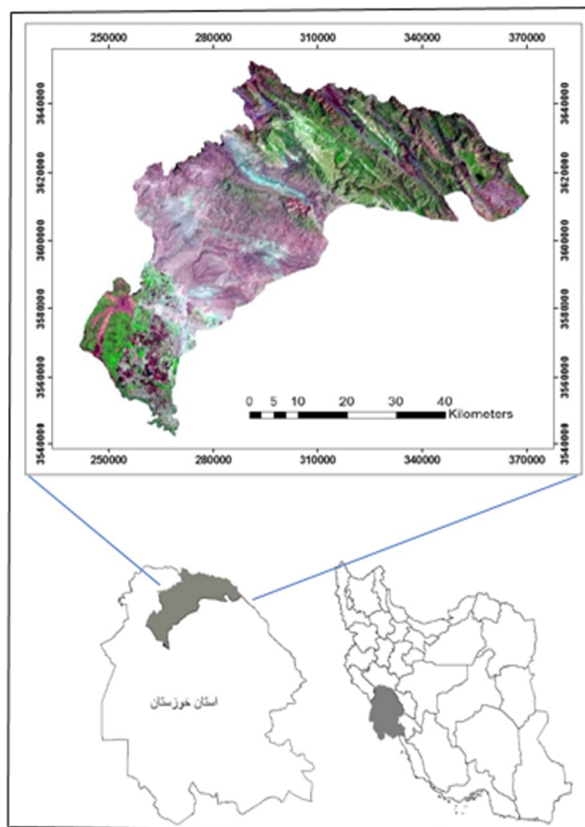
11. *Anchusa strigosa*

بارندگی، دما، ارتفاع، شیب، بافت و عمق خاک استخراج و سپس در ادامه به ارزیابی کارایی مدل‌های داده‌کاوی الگوریتم ازدحام ذرات^۱، شبکه عصبی مصنوعی، رگرسیون لجستیک و نزدیک‌ترین همسایه در تعیین پراکنش گونه گاوزبان خارک‌دار در شهرستان دزفول پرداخته شد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

شهرستان دزفول با ۷۸۸۴ کیلومترمربع وسعت در شمال استان خوزستان واقع گردیده و مرکز آن شهر دزفول است. این منطقه بین طول‌های جغرافیای ۴۸ درجه و ۱۸ دقیقه تا ۴۹ درجه و ۳۳ دقیقه و عرض‌های جغرافیایی ۳۱ درجه و ۵۹ دقیقه تا ۳۲ درجه و ۵۷ دقیقه قرار دارد (شکل ۱). اقلیم شهرستان دزفول نیمه‌خشک تا مدیترانه‌ای بوده که متوسط بارندگی سالانه این شهرستان ۴۰۰ میلی‌متر و متوسط دمای سالانه آن ۳ درجه سانتی‌گراد است و دارای متوسط ارتفاع از سطح دریا ۱۴۳ متر می‌باشد.



شکل (۱): نقشه موقعیت شهرستان دزفول در ایران و استان خوزستان

۲-۲- گونه گیاهی مورد مطالعه

Anchusa strigosa متعلق به تیره گل گاوزبان^۲ با نام فارسی گاوزبان خارک‌دار و نام محلی گاوزبان است (صادقی‌نیا و همکاران، ۱۳۹۵). جنس‌های مهم خانواده گاوزبانیان شامل جنس‌های آفتاب‌پرست^۳ (متجاوز از ۲۰۰ گونه)، سگ‌زبان^۴ (۷۵ گونه)، گاوزبان^۵ (۵۰ گونه)، گاوزبان اروپایی^۶ (۴ گونه) و جنس *Tournefortia* (۱۵۰ گونه) است که این گونه یکی از ۵۰ گونه جنس گاوزبان است (آخوندعلی و بهداروند، ۱۳۹۰؛ مظفریان، ۱۳۸۸). این گیاه به صورت خودرو می‌روید. تکثیر آن از طریق بذر صورت می‌گیرد. گل‌دهی این گیاه در استان خوزستان از اواسط زمستان تا اواخر بهار بوده و بذر آن پس از بهار قابل برداشت می‌باشد (آخوندعلی و بهداروند، ۱۳۹۰). اندام مورد مصرف این گیاه گل و برگ و مواد مؤثره آن آلکالوئیدها و پلی‌فنول است. این گیاه مقوی قلب و آرام‌بخش است (هویزه و همکاران، ۱۳۸۱). کاربردهای این گیاه در طب اسلامی در درمان بیماری‌های پوستی، تب و بیماری‌های تنفسی است (Saad et al., 2005). همچنین از ریشه این گیاهان نوعی ترکیب رزینی با خاصیت رنگی به دست می‌آید (Judd et al., 1999).

1. Particle Swarm Optimization
2. Boraginaceae
3. Heliotropium

4. Cynoglossum
5. Anchusa
6. Borago



شکل (۲): گونه گاوزبان خارک‌دار در دره تابیران دزفول

۲-۳- آماده‌سازی داده‌ها

در این پژوهش برای بررسی شرایط بوم‌شناختی مکان‌های حضور گونه گاوزبان خارک‌دار، شش عامل محیطی از بین عوامل تاثیرگذار بر رشد این گونه که به طور بالقوه می‌توانستند بر گسترش این گونه موثر باشند، بررسی گردید. شاخص‌های بوم‌شناختی انتخاب شده شامل عوامل اقلیمی (بارندگی و دما)، عوامل خاکی (بافت و عمق خاک) و عوامل توپوگرافی (شیب و ارتفاع) می‌باشند. نقشه عوامل اقلیمی با استفاده از داده‌های ایستگاه‌های هواشناسی منطقه و نقشه عوامل خاکی بر مبنای نمونه‌برداری‌های انجام شده از خاک منطقه و با استفاده از روش‌های زمین‌آمار در نرم‌افزار ArcGIS استخراج شد. نقشه عوامل توپوگرافی محدوده نیز بر مبنای مدل رقومی ارتفاعی منطقه (۱۲/۵ متری) دانلود شده از سایت آلاسکا^۱ تهیه شد. برای تهیه لایه‌های اقلیمی با توجه به وسعت زیاد منطقه و کمبود ایستگاه‌های هواشناسی سینوپتیک ابتدا گرادیان ایستگاه‌های داخل و خارج منطقه در محیط اکسل تهیه شد و سپس با استفاده از مدل رقومی ارتفاع^۲ منطقه نقشه همبارش و هم‌دما از متوسط بارندگی و دمای سالانه تهیه شد. نقشه بافت و عمق خاک با توجه به اطلاعات جمع‌آوری شده از نمونه‌برداری‌های صحرایی، در نرم‌افزار ArcGIS تهیه گردید. برای ارزیابی کارایی مدل‌های داده‌کاوی، ابتدا ۳۰ نقطه حضور و ۳۰ نقطه عدم حضور گونه با استفاده از نقشه‌های پوشش گیاهی منطقه که بر مبنای نقاط نمونه‌برداری شده تهیه گردیده بود استخراج گردید و سپس با استفاده از نقشه‌های شاخص‌های اکولوژیکی تهیه شده بر مبنای روش‌های زمین‌آمار در نرم‌افزار ArcGIS مقادیر شاخص‌ها در نقاط حضور و عدم حضور استخراج گردید. به منظور تعیین میزان اثر هر یک از شاخص‌ها در پیش‌بینی حضور و عدم حضور گونه از روش ماشین‌بردار به منظور وزن‌دهی استفاده شد و پراکنش گونه با هر کدام از الگوریتم‌های ازدحام ذرات، رگرسیون لجستیک، شبکه عصبی مصنوعی و نزدیکترین همسایه در محیط نرم‌افزار رپیدمینر^۳ مورد بررسی و مدل‌سازی قرار گرفت. سپس ارزیابی هر یک از روش‌ها با استفاده از معیارهای دقت^۴، همبستگی^۵ و ضریب کاپا انجام و بهترین مدل انتخاب شد.

۲-۴- مدل‌های داده‌کاوی

۲-۴-۱- نزدیک‌ترین همسایه

یک تکنیک دسته‌بندی است که از نسخه‌ای از این روش استفاده می‌کند. در این روش تصمیم‌گیری اینکه یک مورد جدید در کدام دسته قرار گیرد با بررسی تعدادی (K) از شبیه‌ترین موارد یا همسایه‌ها انجام می‌شود. تعداد موارد برای هر کلاس شمرده می‌شوند و مورد جدید به دسته‌ای که تعداد بیشتری از همسایه‌ها به آن تعلق دارند نسبت داده می‌شود (عزمی و عراقی نژاد، ۱۳۹۱).

۲-۴-۲- رگرسیون لجستیک

شبیه رگرسیون چندگانه است که در آن از یک مجموعه متغیر مستقل، توأمًا، برای پیش‌بینی متغیر وابسته استفاده می‌شود. در این روش، متغیر پیش‌بینی کننده در مقیاس کیفی مورد سنجش قرار می‌گیرد. پیش‌بینی‌ها می‌توانند هر نوع ترکیبی از متغیرهای مقوله‌ای را شامل شوند. رگرسیون

1. <https://search.asf.alaska.edu>
2. Discrete Element Method (DEM)
3. RapidMiner

4. Accuracy
5. Correlation

لجستیک زمانی به کار می‌رود که در نظر باشد تا عضویت در یک طبقه پیش‌بینی شود نه مقدار یک متغیر کیفی، یعنی زمانی که متغیر پاسخ (طبقات) فقط دو حالت را به خود می‌گیرد. تکنیک رگرسیون لجستیک، یک تحلیل چند متغیری است که تمامی عوامل پیش‌بینی کننده موجود در یک مساله را به طور همزمان مورد توجه قرار می‌دهد (Rajae, 2011).

۲-۴-۳- شبکه عصبی مصنوعی

در حال حاضر در برنامه‌های مختلفی با موفقیت زیاد در حال استفاده می‌باشند. اولین مزیت اصلی آنها این می‌باشد که نیاز به تعیین الگوریتم حل مشکل توسط خود استفاده کننده را ندارند (همانند مورد برنامه‌ریزی کلاسیک)، بلکه در عوض این شبکه‌ها از مثال‌ها یاد می‌گیرند، درست مثل انسان‌ها. دومین مزیت اصلی آنها این است که آنها دارای توانایی تعمیم ذاتی می‌باشند. بدین معنی که آنها می‌توانند الگوهایی که شبیه هم هستند اما با مواردی که برای آموزش مورد استفاده قرار گرفته‌اند یکسان نمی‌باشند، تشخیص داده و به آنها پاسخ دهند (Benardos and Vosniakos, 2007). شبکه عصبی مصنوعی یکی از مهمترین تکنیک‌های داده‌کاوی می‌باشد.

۲-۴-۴- ازدحام ذرات

یک تکنیک بهینه‌سازی احتمالی است که بر مبنای جمعیت کار می‌کند. این روش توسط Eberhart and Kennedy (۱۹۹۵) ارائه شد و ایده اصلی آن از رفتار دسته جمعی ماهی‌ها یا پرندگان به هنگام جستجوی غذا الهام گرفته شده است. گروهی از پرندگان در فضایی به صورت تصادفی دنبال غذا می‌گردند. تنها یک تکه غذا در فضای مورد بحث وجود دارد. هیچ یک از پرندگان محل غذا را نمی‌دانند. یکی از بهترین استراتژی‌ها می‌تواند دنبال کردن پرنده‌ای باشد که کمترین فاصله را تا غذا داشته باشد. این استراتژی در واقع جانمایه الگوریتم ازدحام ذرات است (امامی میبیدی و همکاران، ۱۳۸۸).

۲-۵- ارزیابی نتایج

هدف از ارزیابی مدل به دست آوردن میزان خطای آن با توجه به داده‌های ورودی به مدل جهت آموزش و براساس معیارهای مختلف محاسبه خطا می‌باشد. در این مطالعه به منظور ارزیابی از معیارهای دقت و ضریب کاپا استفاده شد. دقت، درصد پیش‌بینی‌هایی است که درست صورت گرفته است که از رابطه ۱ به دست می‌آید.

$$\text{رابطه (۱)} = \frac{\text{مجموعه طبقه بندی های صحیح}}{\text{تعداد کل نمونه}} = \text{دقت}$$

ضریب کاپا بیانگر میزان توافق مقادیر مشاهده شده و مقادیر پیش‌بینی شده می‌باشد. مقدار کاپا از صفر تا یک تغییر می‌کند. هر چه مقدار کاپا به یک نزدیک‌تر باشد، نشان‌دهنده توافق بهتر مدل با دنیای واقعی است. ضریب کاپا به عنوان معیار دیگری در بیان صحت، برای هر ماتریس به کمک عناصر قطری و حاشیه‌ای محاسبه شده و نشان‌دهنده آن است که طبقه‌بندی چقدر با داده‌های واقعی توافق دارد. میزان کلی توافق برای هر ماتریس، بر پایه تفاوت بین توافق عملی طبقه‌بندی (توافق بین طبقه‌بندی رایانه‌ای و داده‌های واقعی زمینی که توسط عناصر قطری جدول نمایش داده می‌شوند) و توافق شانسی (که از مقادیر فرعی پیکسل به دست می‌آید) محاسبه می‌گردد.

۲-۶- وزن دهی

تمامی پارامترهای ورودی مدل دارای تاثیر و اهمیت یکسانی در پیش‌بینی نمی‌باشند. برخی از پارامترها دارای همبستگی و ارتباط بیشتری با خروجی مدل بوده و تاثیر بیشتری در پیش‌بینی‌ها داشته است. وزن‌دهی پارامترها یکی از تکنیک‌های داده‌کاوی است. در این مطالعه به منظور وزن‌دهی از الگوریتم ماشین‌بردار پشتیبان استفاده شده، این الگوریتم ضرایب بردار نرمال یک ماشین‌بردار پشتیبان خطی را به عنوان وزن ویژگی تعیین می‌کند (صنعی آبا‌ده و همکاران، ۱۳۹۴؛ Asadi et al, 2021).

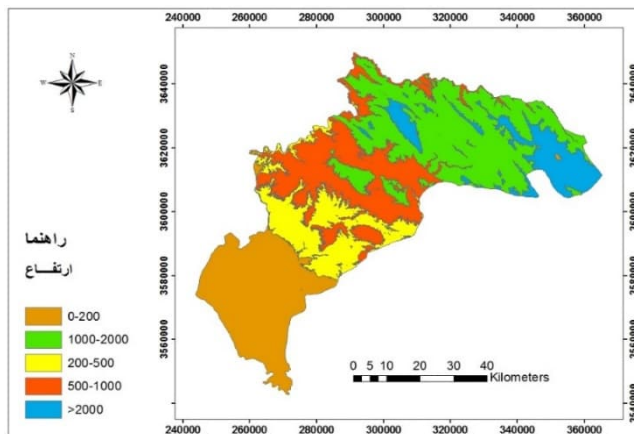
۳- نتایج

اطلاعات آماری نشان می‌دهد که پارامترهای بارندگی و درجه حرارت در مناطق حضور گونه دارای کمترین ضریب تغییرات بوده‌اند و پارامترهای شیب، ارتفاع و عمق خاک دارای بیشترین ضریب تغییرات بوده است و جدول ۱ خلاصه آماری پارامترهای بیوفیزیکی مربوط به مناطق حضور و عدم حضور را نشان می‌دهد.

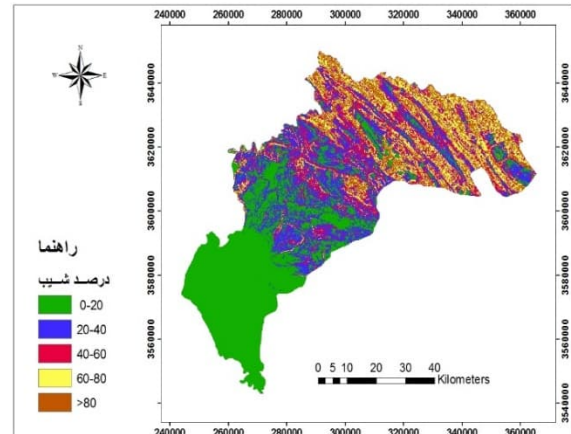
جدول (۱): اطلاعات آماری پارامترها در مناطق حضور و عدم حضور گونه

منطقه	پارامتر	حداقل	حداکثر	میانگین	انحراف معیار	ضریب تغییرات
حضور گونه	شیب	۰	۴۰	۲۰	۲۰	۱
	ارتفاع	۲۰۰	۱۰۰۰	۶۰۰	۴۰۰	۰/۷
	بارندگی	۲۵۱	۴۰۰	۳۲۵/۵	۷۴/۵	۰/۲
	درجه حرارت	۱۸	۲۷	۲۲/۵	۴/۵	۰/۲
	بافت خاک	ریگ درشت	لومی سیلتی	-	-	-
	عمق خاک	۳	۴۰	۲۱/۵	۱۸/۵	۰/۹
عدم حضور گونه	شیب	۰	۱۰۰	۵۰	۵۰	۱
	ارتفاع	۴۵	۲۱۰۰	۱۰۵۰	۱۰۵۰	۱
	بارندگی	۲۵۱	۹۰۰	۵۷۵/۵	۳۳۴/۵	۰/۶
	درجه حرارت	۰	۲۷	۱۳/۵	۱۳/۵	۱
	بافت خاک	ریگ درشت	لومی سیلتی	-	-	-
	عمق خاک	۰/۲	۴۰	۴۰	۲۰	۰/۵

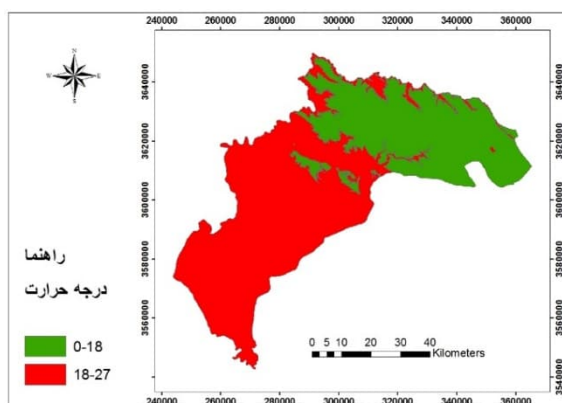
بررسی اطلاعات آماری پارامترها در دو منطقه نشان داد که در مناطق حضور گونه نسبت به مناطق عدم حضور گونه میانگین بارندگی کمتر و درجه حرارت بالاتر می‌باشد. مناطق حضور گونه از شیب و ارتفاع کمتری نسبت به مناطق عدم حضور گونه برخوردار می‌باشد. در شکل‌های ۲ تا ۷ نقشه شیب، ارتفاع، بارندگی، درجه حرارت، بافت و عمق خاک که در نرم‌افزار ArcGIS تهیه شده، آمده است.



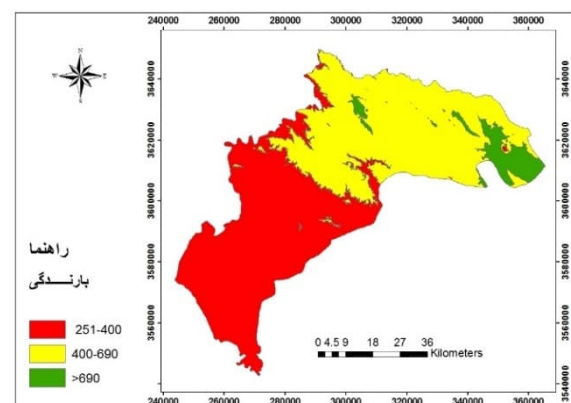
شکل (۳): نقشه ارتفاع منطقه



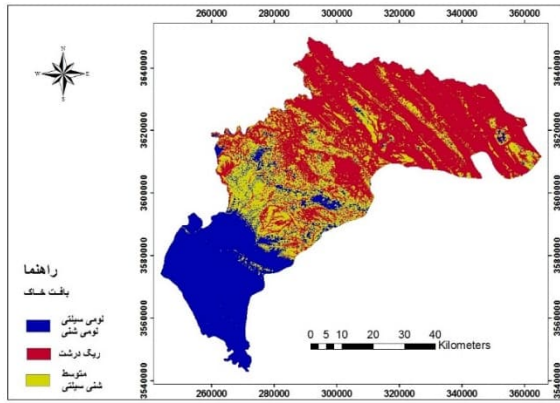
شکل (۲): نقشه شیب منطقه



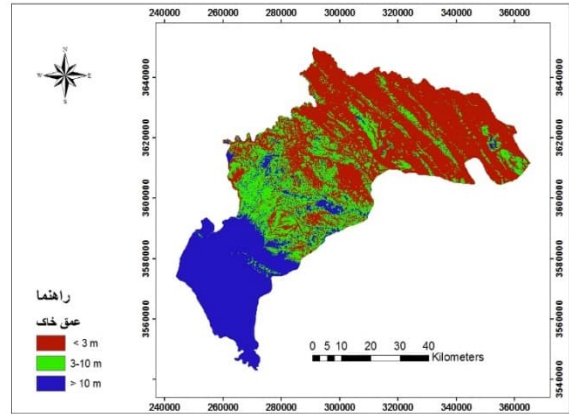
شکل (۵): نقشه میانگین دمای منطقه



شکل (۴): نقشه بارندگی منطقه



شکل (۷): نقشه بافت خاک منطقه

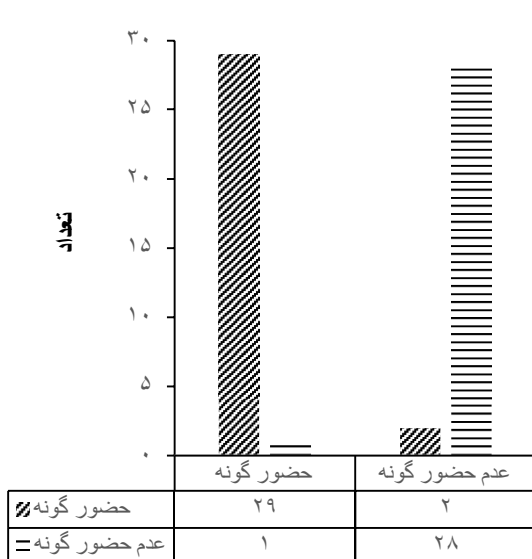


شکل (۶): نقشه عمق خاک منطقه

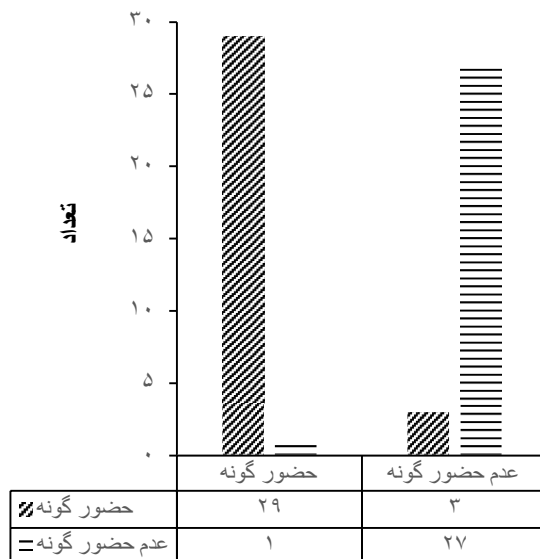
باتعیین مناطق پراکنش گونه و استخراج پارامترهای مورد نظر در مناطق حضور و عدم حضور گونه گیاهی گاوزبان، داده‌ها با فرمت مناسب به نرم‌افزار رپیدماینر وارد و پس از تنظیم پارامترهای مدل، رفتارهای هر یک از مدل‌ها بر روی مجموعه داده معرفی شده مورد بررسی و ارزیابی قرار گرفت. نتایج به دست آمده از این الگوریتم‌ها نیز مورد بررسی و ارزیابی قرار گرفت و میزان صحت آنها سنجیده شد. جدول ۲ نتایج ارزیابی صحت نتایج به دست آمده را با مقادیر دقت و ضریب کاپا نشان می‌دهد. نتایج نشان داد که به طور کلی هر چهار الگوریتم داده‌کاوی دارای دقت بالایی در پیش‌بینی مکان حضور گونه گاوزبان خارک‌دار بوده است. مقایسه میزان دقت روش‌های مختلف نشان داد که الگوریتم‌های ازدحام ذرات و شبکه عصبی مصنوعی با دقت ۰/۹۵ و ضریب کاپا ۰/۹ دارای بیشترین دقت در پیش‌بینی رویشگاه بوده است. نتایج مدل شبکه عصبی مصنوعی نشان داد از ۳۰ نقطه مورد ارزیابی حضور گونه تنها مدل قادر به تشخیص یک نقطه نبوده و در مناطق عدم حضور مدل قادر به تفکیک ۳ نقطه عدم حضور نبوده است. مقایسه نقاط حضور و عدم حضور که در طبقه صحیح یا ناصحیح واقع شده‌اند برای سایر مدل‌ها در شکل‌های ۸ تا ۱۱ ارائه شده است.

جدول (۲): نتایج ارزیابی صحت نتایج با مقادیر دقت و کاپا

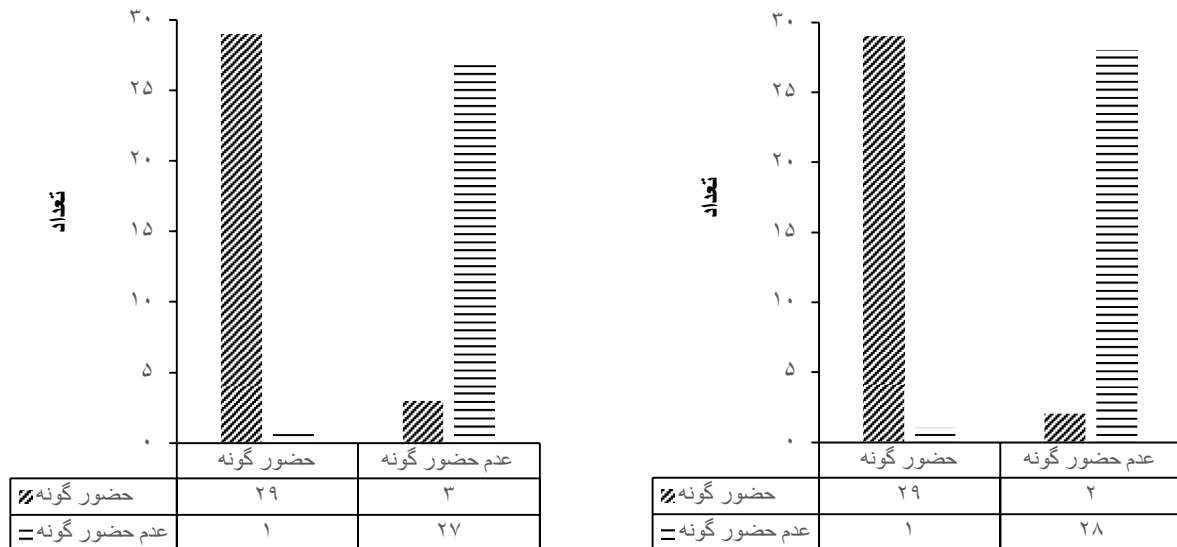
مدل	دقت	ضریب کاپا	همبستگی
ازدحام ذرات	۹۵٪	۰/۹	۰/۹۰۱
رگرسیون لجستیک	۹۳/۳۳٪	۰/۸۶۷	۰/۸۶۷
شبکه عصبی مصنوعی	۹۵٪	۰/۹	۰/۹۰۱
نزدیک‌ترین همسایه	۹۳/۳۳٪	۰/۸۶۷	۰/۸۶۹



شکل (۹): وضعیت پیش‌بینی حضور و عدم حضور با استفاده از مدل رگرسیون لجستیک

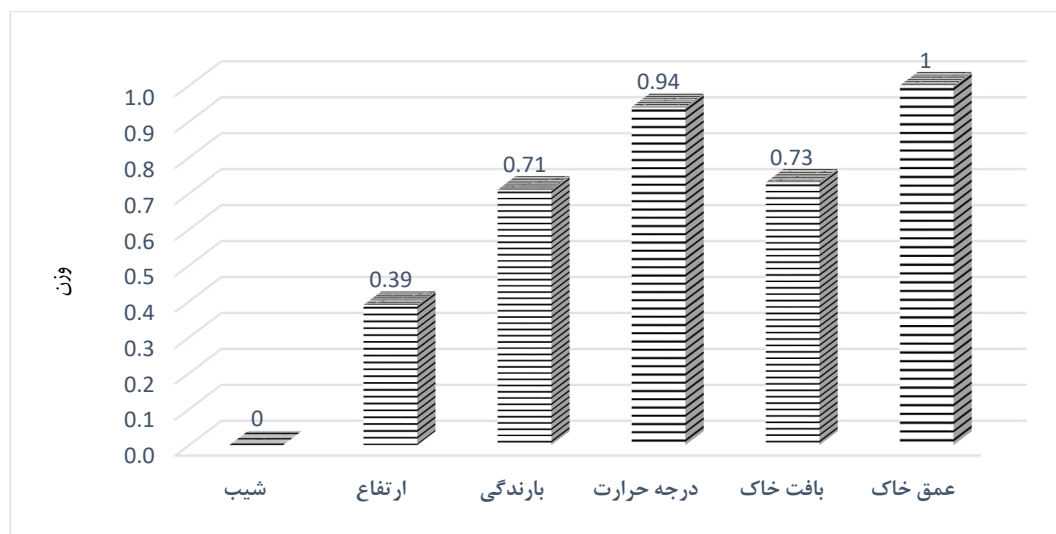


شکل (۸): وضعیت پیش‌بینی حضور و عدم حضور با استفاده از مدل شبکه عصبی مصنوعی



شکل (۱۰): وضعیت پیش‌بینی حضور و عدم حضور با استفاده از مدل ازدحام ذرات
شکل (۱۱): وضعیت پیش‌بینی حضور و عدم حضور با استفاده از مدل نزدیک‌ترین همسایه

با توجه به تاثیر متفاوت هر یک از پارامترها در تعیین محل مناسب رویش گونه وزن‌دهی بین پارامترها بر اساس روش ماشین بردار پشتیبان انجام شد، نتایج حاصل از وزن‌دهی در قالب نمودار شکل ۱۲ نشان داده شده است. نتایج نشان داد پارامتر عمق خاک با ضریب ۱ و پارامتر درجه حرارت با ضریب ۰/۹۴ دارای بیشترین تاثیر در تعیین رویشگاه‌های مرتعی گونه گاوزبان خاک‌دار داشته است. نتایج همچنین نشان داد پارامتر شیب با ضریب تاثیر صفر و پارامتر ارتفاع با ضریب تاثیر ۰/۳۹ دارای کمترین ضریب تاثیر در مدل‌سازی تعیین رویشگاه مرتعی گونه گاوزبان خاک‌دار داشته است.



شکل (۱۲): نمودار نتایج وزن پارامترهای مختلف بیوفیزیکی در حضور یا عدم حضور گونه مورد نظر

۴- بحث و نتیجه‌گیری

به طور معمول مدل‌های پراکنش گونه‌ای، از روابط بین متغیرهای محیطی و رخداد‌های گونه‌ها برای تعریف شرایط محیطی که جمعیت می‌تواند در آن باقی بماند استفاده می‌نمایند و به صورت یک نقشه پراکنش مکانی محیط‌هایی که مناسب برای گونه‌ها هستند را نشان می‌دهد. در مدل‌های پراکنش گونه‌ای فرض بر این است که می‌تواند پراکنش مکانی بالقوه یک گونه گیاهی را به وسیله داده‌های حضور و عدم حضور متغیر پاسخ و متغیرهای پیش‌بینی کننده (پیشگو) تعیین کنند. معمول‌ترین روش برای پیش‌بینی توزیع جغرافیایی بالقوه یا واقعی یک گونه استفاده از شرایط محیطی است که برای رشد گونه مناسب است این شرایط می‌تواند به وسیله یکی از روش‌های مدل‌سازی مکانی یا آماری تعیین شود. مدل‌های پراکنش گونه‌ای ابزار مناسبی برای پیش‌بینی پتانسیل پراکنش گونه‌ها در ارتباط با متغیرهای محیطی هستند. یکی از اهداف این مطالعه بررسی توانایی مدل‌های تجربی در پیش‌بینی مناطق حضور بالقوه گونه گیاهی است.

نتایج این مطالعه نشان داد که مهمترین عامل محیطی اثرگذار در پراکنش این گونه گیاهی، شاخص بارندگی می‌باشد، نتایج مطالعه Fisher et al. (۱۹۸۸)، نیز نشان داد میزان رطوبت قابل دسترس از مهمترین عوامل موثر بر استقرار گیاهان می‌باشد، که با نتایج مطالعه انجام شده همخوانی دارد. نتایج مطالعه حاکی از این مهم می‌باشد که فاکتورهای محیطی تاثیر مشابه در پراکنش گونه‌های مرتعی نداشته و بعضی فاکتورها را می‌توان به عنوان عامل معرف نام برد، در صورتی که برخی دیگر از فاکتورها ممکن است هیچ گونه تاثیری بر پراکنش گونه مورد نظر نداشته باشند. نتیجه این تحقیق با نتایج مطالعات محمدی و کاظمی (۱۳۸۶)، و علی اکبری و همکاران (۱۳۸۹) مطابقت دارد.

نظر به اینکه در ایران مطالعاتی در زمینه مطلوبیت رویشگاهی گونه‌های با ارزش مرتعی برای مدیریت و حفاظت از گونه‌های خوشخوارک و در معرض خطر انقراض و تهیه نقشه‌های رویشگاهی با این روش کمتر صورت گرفته است، چنین مطالعاتی می‌تواند الگویی برای کاربرد روش‌های نوین مدیریتی به منظور تعیین مطلوبیت رویشگاه گونه‌های مناسب و برای اصلاح مراتع و حفاظت گونه‌های با ارزش دارویی و صنعتی باشد. نتایج اعمال مدل‌ها بر روی داده‌ها نشان می‌دهد که مدل‌های داده‌کاوی با استفاده از پارامترهای موجود به خوبی قادر به تعیین رویشگاه‌های گونه‌ها می‌باشند. همچنین نتایج حاکی از آن می‌باشد که الگوریتم ازدحام ذرات، شبکه عصبی مصنوعی و نزدیک‌ترین همسایه دارای دقت بیشتری در تعیین رویشگاه می‌باشند. این نتایج با نتایج زارع چاهوکی (۱۳۹۳) و Zare Chahouki and Zare Chahouki (۲۰۱۰) مطابقت داشته ولی با نتایج Manel et al. (۱۹۹۹) همخوانی ندارد.

منابع

- آخوندعلی، ع.ر.، و بهداروند، م. (۱۳۹۰). طرح احیا و بهره‌برداری *Anchusa strigosa* در دزفول. اداره منابع طبیعی شهرستان دزفول.
- اسدی، م.، خاتم‌ساز، م.، و معصومی، ا. (۱۳۸۱). کتاب فلور ایران (جلد ۳۹، تیره *Boraginaceae*). موسسه تحقیقات جنگل‌ها و مراتع، ۵۰۸.
- تقی‌زاده مهرجردی، ر.ا.، سرمیدیان، ف.، امید، م.، تومانیان، ن.، روستا، م.ج. و رحیمیان، م.ح. (۱۳۹۳). نقشه برداری رقمی کلاس‌های خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه اردکان استان یزد. مهندسی زراعی، (۳۷(۲)، ۱۱۵-۱۰۱.
- امامی‌میبیدی، ع.، خضری، م.، و عظیمی، ا. (۱۳۸۸). شبیه‌سازی تابع تقاضای انرژی در ایران با استفاده از الگوریتم بهینه سازی انبوه ذرات (PSO). مطالعات اقتصاد انرژی، ۲۰، ۱۵۹-۱۴۱.
- پیری صحراگرد، ح.، زارع چاهوکی، م.ع.، و آذرینوند، ح. (۱۳۹۴). مدل‌سازی پراکنش گونه‌های گیاهی مناطق خشک و بیابانی با استفاده از روش شبکه‌های عصبی مصنوعی (بررسی موردی: مراتع حوض سلطان استان قم)، مدیریت بیابان، (۳(۵)، ۳۹-۲۶.
- پیری صحراگرد، ح. (۱۳۹۶). مدل‌سازی پیش‌بینی پراکنش رویشگاه گونه‌های گیاهی با روش رگرسیون لجستیک (مطالعه موردی: مراتع غرب تفتان، شهرستان خاش). پژوهش‌های گیاهی، (۳۰(۴)، ۸۰۶-۷۹۲.
- دهقانی‌نژاد، س.، علوی، س.ج.، و حسینی، س.م. (۱۳۹۷). قابلیت پیش‌بینی مدل‌های آماری در ارزیابی توان تولید رویشگاه راش شرقی. اکوسیستم‌های طبیعی ایران. (۲(۹)، ۳۰-۱۳.
- زارع چاهوکی، م.ع. (۱۳۸۵). مدل‌سازی پراکنش گونه‌های گیاهی مراتع مناطق خشک و نیمه‌خشک (مطالعه موردی: مراتع پشتکوه استان یزد). رساله دکتری مرتعداری، دانشکده منابع طبیعی دانشگاه تهران، کرج، ۱۸۰ص.
- زارع چاهوکی، م.ع.، خلاصی اهوازی، ل.، و آذرینوند، ح. (۱۳۹۳). مدل‌سازی پراکنش گونه‌های گیاهی بر اساس عوامل خاک و توپوگرافی با استفاده از روش رگرسیون لجستیک در مراتع شرق سمنان. مرتع و آبخیزداری، (۱(۶۷)، ۵۹-۴۵.
- فاطمی، س.س.، رحیمی، م.، ترکش، م.، و روانبخش، ه. (۱۳۹۶). پیش‌بینی پراکنش جغرافیایی گونه *Juniperus excelsa* M.Bieb. با استفاده از داده‌های اقلیمی در شرایط حاضر و آینده در استان سمنان. جنگل ایران، (۲(۹)، ۲۴۸-۲۳۳.
- صادقی‌نیا، م.، تازه، م.، جعفری، ز.، و کیانی، ک. (۱۳۹۵). تعیین رویشگاه بالقوه گاوزبان خارکدار (*Anchusa strigosa*) با استفاده از تحلیل سلسله مراتبی و سیستم اطلاعات جغرافیایی در شهرستان دزفول. سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی، (۴(۷)، ۳۰-۱۸.
- صنعی‌آباد، م.، محمودی، س.، و طاهرپور، م. (۱۳۹۴). داده‌کاوی کاربردی (ویراست دوم). تهران: انتشارات نیاز دانش، ۵۳۶ص.
- عزمی، م.، و عراقی‌نژاد، ش. (۱۳۹۹). توسعه روش رگرسیون و K نزدیکترین همسایگی در پیش‌بینی جریان رودخانه. آب و فاضلاب، ۲، ۱۱۸-۱۰۷.
- علی‌اکبری، م.، جعفری، ر.، وهابی، م.ر.، و سعادت‌فرا، ا. (۱۳۸۹). تعیین رویشگاه بالقوه گونه‌گون زرد با استفاده از تلفیق GIS و سنجش‌ازدور. سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی، (۱(۱)، ۳۰-۱۵.
- محمدی، ح.، و کاظمی، م. (۱۳۸۶). تعیین اراضی مناسب کشت بادام دیم در استان آذربایجان شرقی با استفاده از GIS. زراعت و باغبانی، ۴.
- مظفریان، و.ا. (۱۳۸۸). فرهنگ نام‌های گیاهان ایران. تهران: انتشارات فرهنگ معاصر، ۷۴۰ص.
- Asadi, M., Fathzadeh, A., Kerry, R., Ebrahimi-Khusfi, Z., and Taghizadeh-Mehrjardi, R. (2021). Prediction of river suspended sediment load using machine learning models and geo-morphometric parameters. *Arabian Journal of Geosciences*, 14(18), 1-14.
- Benardos, P.G., Vosniakos, G.C. (2007). Optimizing feedforward Artificial Neural Network Architecture. *Engineering Applications of Artificial Intelligence*, 20(3), 365-382.
- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- Eberhart, R., and Kennedy, J. (1995). A new optimizer using particle swarm theory. October 1995.
- Fisher, F.M., Zak, J.C., Cunningham, G., and Whitford, W.G. (1988). Water and nitrogen effect pattern of creosote bush in northern Chihuahuan Desert. *Journal of Range Management*, 41(5), 384-391.

- Franklin, J. (1995). Predictive Vegetation Mapping: Geographic Modeling of Bio spatial Patterns in Relation to Environmental Gradients. *Progress in Physical Geography*, 19(4), 474-499. Proceedings of the Sixth International Symposium on Micro Machine and Human Science
- Frawley, W.J., Piatetsky-Shapiro, G., and Matheus, C.J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine*, 13(3), 57-70.
- Guisan, A., and Zimmermann, N. (2000). Predictive Habitat Distribution Models in Ecology. *Ecological Modeling*, 135(2-3), 147-186.
- Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F. and Donoghue, M.J. (1999) *Plant Systematics: A Phylogenetic Approach*. Sinauer Associates is an imprint of Oxford University, 677p.
- Manel, S., Dias, J.M., and Ormerod, S.J. (1999). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: A case study with a Himalayan river bird. *Ecological Modeling*, 120(2-3), 337-347.
- Rajae, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. *Science of The Total Environment*, 409(15), 2917-2928.
- Saad, B., Azaizeh, H., and Said, O. (2005). Tradition and Perspectives of Arab Herbal Medicine: A review. *Evidence-Based Complementary and Alternative Medicine*, 4(2), 475-479.
- Zare Chahouki, M.A., Zare Chahouki, A. (2010). Predicting the distribution of plant species using logistic regression (Case study: Garizat rangelands of Yazd province). *Desert*, 15(2), 151-158.

Evaluating the effectiveness of data mining models in determining the distribution of the plant species *Anchusa strigosa* (case study: Dezful)

Maryam Asadi *¹, Zahra Jafari ²



Research Article

1. Ph.D. in Watershed Management, Department of Natural Resources and Watershed Management of Khansar, Khansar, Iran.

m.asadi.96@ut.ac.ir

* Corresponding author

2. M.Sc. in Range Management, Department of Nature Engineering Faculty of Agriculture and Natural Resources, Ardakan University, Iran.

zahra.jafari16@yahoo.com

Article Code: 2206-1027
Continous Pagation: 234-245

Received: 15 June 2022
Accepted: 15 January 2023
Online: 22 January 2022
Review speed: 214 days

Citation:

Asadi, M., and Jafari, Z. (2022). Evaluating the effectiveness of data mining models in determining the distribution of the plant species *Anchusa strigosa* (case study: Dezful). *Management of Natural Ecosystems*, 2(3), 24-35.

Abstract

Industrial and medicinal plants are one of the by-products of rangelands, the exploitation of which has a role in the economy of rangelands. The preservation and development of distribution of medicinal plants at the level of rangelands is important from various economic, social and technical aspects of rangelands. There are many industrial and medicinal plant species in the rangeland. *Anchusa strigosa* is a medicinal and indigenous species of Dezful. One of the important issues in the principled exploitation of these medicinal plants is to determine, its habitat areas. The purpose of this research is to investigate the relationship between the biophysical properties of the potential habitat of this species using different data mining techniques. Evaluating the effectiveness of data mining models, first 30 presence points and 30 absence points of the species were recorded using vegetation map and field studies. Then the parameters of height, slope, temperature, rainfall, texture and depth of soil were extracted from the obtained maps using geostatistical methods based on regional data. Determination of the influence coefficient of each of the mentioned parameters was done using the vector machine method. Then, the presence of the mentioned species was predicted based on biophysical parameters using data mining methods. The data mining algorithms used in this study include particle swarm, logistic regression, artificial neural networks and the nearest neighbor. The results showed that particle swarm algorithm and artificial neural network with accuracy rate of 0.95 and kappa coefficient of 0.9 have a higher accuracy in determining the species presence areas. The results of this study showed that using data mining methods, it is possible to predict the presence of the desired species, based on biophysical parameters with high accuracy.

Key Words: Particle Swarm Algorithm, Habitat Prediction, Logistic Regression, Artificial Neural Networks, Nearest Neighbor.