

کاربرد مدل M5 در برآورد تبخیر- تعرق مرجع در ایستگاه‌های

شیراز و کرمانشاه

مهسا سامتی، نوذر قهرمان^{۱*} و خلیل قربانی

دانشجوی کارشناسی ارشد گروه مهندسی آبیاری و آبادانی دانشگاه تهران؛ mahsasameti@ut.ac.ir

دانشیار گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران؛ nghahreman@ut.ac.ir

استادیار گروه مهندسی آب دانشگاه کشاورزی و منابع طبیعی گرگان؛ ghorbani.khalil@yahoo.com

چکیده

در این تحقیق توانایی مدل درختی M5 برای تخمین تبخیر- تعرق مرجع در دو ایستگاه هواشناسی شیراز و کرمانشاه مورد ارزیابی قرار گرفت. داده‌های مورد استفاده شامل متغیرهای متوسط دمای هوا، ساعت آفتابی، بارش، دمای نقطه شبنم، رطوبت نسبی، سرعت باد و فشار بخار واقعی روزانه بود. در این تحقیق دو روش پنمن-مانتیت و هارگرویز-سامانی به منظور مقایسه انتخاب و عملکرد مدل درختی M5 به عنوان روشی جدید برای برآورد تبخیر-تعرق بررسی شد. نتایج نشان داد که در هر دو ایستگاه، مدل درختی M5 در تخمین تبخیر-تعرق پتانسیل نسبت به هر دو روش دارای عملکرد مناسبی بود اما توانایی آن در تخمین تبخیر-تعرق به روش پنمن-مانتیت (ایستگاه شیراز: $R^2=0.346$ ، $RMSE=0.975$ ، $RMSE=0.844$)، ایستگاه کرمانشاه: $R^2=0.773$ ، $RMSE=0.362$) نسبت به روش هارگرویز-سامانی (ایستگاه شیراز: $R^2=0.837$ ، $RMSE=0.844$)، ایستگاه کرمانشاه: $R^2=0.774$ ، $RMSE=0.774$) بالاتر شد. آنالیز حساسیت نشان داد در روش پنمن-مانتیت در ایستگاه شیراز متغیرهای متوسط دمای هوا، ساعت آفتابی، دمای نقطه شبنم، متوسط سرعت باد و در ایستگاه کرمانشاه دمای هوا، ساعت آفتابی، متوسط سرعت باد، رطوبت نسبی و دمای نقطه شبنم، بیشترین تاثیر را داشتند.

واژه های کلیدی: آنالیز حساسیت، پنمن-مانتیت، داده کاوی، مدل درختی، هارگرویز-سامانی

مقدمه

شرایط اقلیمی مختلف ارزیابی شده است (آلن و همکاران، ۱۹۹۸؛ کاتول و همکاران، ۱۹۹۲؛ آماتیا و همکاران، ۱۹۹۵؛ اسمیت و همکاران، ۱۹۹۶؛ ونتورا و همکاران، ۱۹۹۹؛ برنجینا و گاویلان، ۲۰۰۵). نتایج تحقیقات نشان داد که معادله پنمن-مانتیت به عنوان یک معادله استاندارد جهانی برای تخمین ET_0 برتر شد (آلن و همکاران، ۱۹۹۸). در معادله پنمن-مانتیت به

روش استاندارد مورد استفاده برای تعیین تبخیر-تعرق گیاه مرجع، روش لایسیمتری است که در صورت عدم دسترسی، از روش‌های تجربی استفاده می‌شود. برخی روش‌های تجربی از اصول فیزیکی می‌کنند و بعضی از روش‌های آماری و از همبستگی آماری بین تبخیر-تعرق و چند متغیر محدود استفاده می‌کنند (شارما، ۱۹۸۵). عملکرد روش‌های تجربی در شرایط اقلیمی مختلف ارزیابی شده است (آلن و

^۱ آدرس نویسنده مسؤول: کرج، پردیس کشاورزی و منابع طبیعی، گروه مهندسی آبیاری و آبادانی کدپستی: ۳۱۵۸۷۷۷۷۷۱-صندوق

پستی: ۷۷۸۷۱

* دریافت: بهمن، ۱۳۹۰ و پذیرش: اردیبهشت، ۱۳۹۲

داده‌های دمای بیشینه و کمینه هوا، رطوبت بیشینه و کمینه (یا فشاربخار واقعی)، سرعت باد در ارتفاع ۲ متری و تابش خورشید (یا ساعات آفتابی) نیاز است. درحالی‌که داده دمای هوا در بیشتر ایستگاه‌های هواشناسی سراسر جهان قابل دسترس است، سایر متغیرها همیشه در دسترس نیستند (دروگوز و آلن، ۲۰۰۲). از طرفی خطاهای تصادفی و یا سیستماتیک در متغیرهای هواشناسی در تخمین ET_0 ، باعث خطاهای مشهودی در برآورد می‌شود (میر و همکاران، ۱۹۸۹). در این شرایط، گزینه دیگر استفاده از روش‌های تخمین ET_0 است که فقط به داده دما اکتفا می‌کنند. لیکن در چنین شرایطی معادله‌ی تجربی (هارگریوز و سامانی، ۱۹۸۲) به عنوان یکی از روش‌های مناسب برای مناطق با داده کم که عمدتاً داده روزانه دما موجود باشد، پیشنهاد شد. معادله‌ی هارگریوز-سامانی فقط به داده دمای بیشینه و کمینه و تابش برون-زمینی نیاز دارد و چون تابش برون‌زمینی می‌تواند بصورت تئوری محاسبه شود (دروگوز و آلن، ۲۰۰۲)، تنها متغیر مورد نیاز، دمای هوا است. تبخیر-تعرق به روش هارگریوز-سامانی را می‌توان بصورت روزانه محاسبه کرد، هر چند در مقیاس هفتگی و ماهیانه نتایج بهتری بدست آمد (جنسن و همکاران، ۱۹۹۰؛ چویسنل و همکاران، ۱۹۹۲؛ دروگوز و آلن، ۲۰۰۲؛ هارگریوز و آلن، ۲۰۰۳).

در سال‌های اخیر از تکنیک‌های بسیاری به منظور برآورد متغیرهای هواشناسی استفاده شد که اغلب این تکنیک‌ها داده محور بوده و سعی نمودند با داشتن اطلاعاتی در مورد روابط بین داده‌ها در یک سری زمانی بلند مدت به تشخیص الگوی آن متغیر پرداخته تا بتوانند به نحوی متغیر مورد نظر را که دسترسی به آن در برخی مناطق میسر نیست برآورد نمایند. از جمله آن‌ها می‌توان به شبکه‌های عصبی مصنوعی اشاره نمود. معمولاً شبکه‌های عصبی آموزش داده می‌شوند تا یک ورودی خاص به هدف خروجی خاصی برسد. شبکه براساس مقایسه مقدار خروجی و هدف مقایسه می‌شود تا زمانی‌که جمع مربع اختلافات بین مقادیر هدف و خروجی حداقل گردد.

رهیافت دیگری که اخیراً بر اساس تکنیک‌های داده‌کاوی برای تخمین تبخیر-تعرق مرجع ارائه شده است، مدل درختی M5 بود که در مطالعات منابع آب به کارگرفته شد. بیشتر کاربردهای مدل درختی M5 در منابع علمی عمدتاً ناظر بر پیش‌بینی است، اما این روش برای مسائل طبقه‌بندی نیز عملکرد رضایت بخشی داشت (پال، ۲۰۰۶).

داده‌کاوی (DM^1) اغلب بصورت فرآیند استخراج اطلاعات معتبر، از پیش تعیین شده و قابل فهم برای پایگاه‌های بزرگ داده تعریف می‌شود تا تصمیمات تجاری را بهبود داده و بهینه نماید (براه و شیلویسی، ۲۰۰۲) در سایر تعاریف، داده‌کاوی بصورت تشخیص ساختار در داده است که این ساختار، الگوها، مدل‌های آماری یا پیش‌بینی داده و روابط بین بخش‌های داده را طراحی می‌نماید (فیاد و اتوروسامی، ۲۰۰۲). داده-کاوی در طیف وسیعی از رشته‌های علوم اجتماعی و زیست محیطی برای پیش‌بینی بکارگرفته شده و طی سال‌های اخیر داده‌کاوی در تحلیل سایر انواع داده‌های علمی مانند بیوانفورماتیک، نجوم و پزشکی نیز تجربه شد (لی و لی شو، ۲۰۰۴).

ترزوی و همکاران (۲۰۰۵) از روش‌ها داده‌کاوی برای مدل‌سازی تبخیراز تشت استفاده نمودند و با بکارگیری الگوریتم ژنتیک دمای هوا، دمای آب و رطوبت نسبی را بعنوان موثرترین عوامل دمایی گزارش نمودند. آنها با استفاده از الگوریتم‌های Kstar، M5Rules و جدول تصمیم مدل‌هایی را توسعه داده و نشان دادند که مدل Kstar در مقایسه با سایر مدل‌ها نتایج بهتری داشت. ترزوی در سال ۲۰۰۷ با بکارگیری داده‌های هواشناسی دریاچه‌ی اگیردیر واقع در ترکیه متغیرهای دمای هوا، دمای آب، تابش خورشید و رطوبت نسبی را به عنوان عوامل مؤثر بر تبخیر از تشت تعیین نمود. در این مطالعه، از الگوریتم-های داده‌کاوی جدول تصمیم، Kstar، M5، M5 Rules، شبکه عصبی و روش‌های رگرسیونی استفاده شد و مقایسه

^۱ Data Mining

نرم افزار داده‌کاوی وکا^۱ نسخه ۳.۶.۴ استفاده شده است. این نرم‌افزار، یک واسط همگون برای بسیاری از الگوریتم‌های یادگیری متفاوت، فراهم کرده است که از طریق آن روش‌های پیش‌پردازش، پس‌پردازش و ارزیابی نتایج طرح‌های یادگیری روی همه مجموعه‌های داده موجود، قابل اعمال است. نرم افزار وکا، پیاده سازی الگوریتم‌های مختلف یادگیری را فراهم می‌کند و به آسانی می‌توان آنها را به مجموعه داده خود اعمال کرد.

برای تخمین تبخیر-تعرق مرجع، پنج متغیر ورودی شامل متوسط دما، ساعات آفتابی، متوسط رطوبت نسبی، متوسط سرعت باد و فشاربخار واقعی استفاده شد. ضمناً هر گونه داده با مقادیر گمشده از سری داده حذف گردید.

لازم به ذکر است در صورت دسترسی به داده های لایسیمیتری معتبر و منتشر شده، این داده ها به عنوان مبنای مقایسه مورد استفاده قرار می‌گرفت که قطعاً بر دقت تحلیل‌ها می‌افزود.

معادلات ۱ و ۲ به ترتیب مربوط به روشهای پنمن-ماتنیث فائو ۵۶ و هارگریوز-سامانی می باشد.

(۱)

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{890}{T_a + 273} U_2 [e_s - e_a]}{\Delta + \gamma(1 + 0.34U)}$$

که ET_0 ، تبخیرتعرق گیاه مرجع ($mm.day^{-1}$)، R_n ، تابش خالص روزانه ($MJ.m^{-2}.day^{-1}$)، G ، شار حرارتی روزانه‌ی خاک ($MJ.m^{-2}.day^{-1}$)، T_a ، متوسط دمای هوای روزانه در ارتفاع ۲ متری ($^{\circ}C$)، U_2 ، متوسط سرعت باد در ارتفاع ۲ متری ($m.s^{-1}$)، e_s ، فشاربخار اشباع (kPa)، e_a ، فشاربخار واقعی (kPa)، Δ شیب فشاربخار اشباع در مقابل منحنی دما ($KPa^{\circ}C$)، γ ثابت سایکرومتری ($KPa^{\circ}C$). در این مطالعه Δ ، R_n ، e_s و e_a با استفاده از معادلات ($\alpha = 0.23$) ارائه شده بوسیله آلن و همکاران (۱۹۹۸) محاسبه شدند. شارحرارتی

نتایج آنها با مقادیر اندازه‌گیری شده تحت تبخیر روزانه نشان داد که الگوریتم M5 بهترین عملکرد را در تخمین تبخیر از تحت داشت.

پال و دسوال (۲۰۰۹) از مدل درختی M5 برای مدل‌سازی تبخیر-تعرق مرجع در ایستگاه دیویس کالیفرنیا استفاده کردند و توانایی این مدل را با روش‌ها پنمن-ماتنیث و هارگریوز-سامانی کالیبره شده مقایسه نمودند. نتایج نشان داد که مدل درختی M5 در پیش‌بینی تبخیر-تعرق مرجع در مقایسه با پنمن-ماتنیث و هارگریوز سامانی کالیبره شده همبستگی خوبی دارد. علاوه بر این، مهم‌ترین متغیرهای موثر بر پیش‌بینی تبخیرتعرق مرجع با استفاده از این مدل شامل تابش خورشید، متوسط دما، رطوبت نسبی و سرعت باد بود.

پژوهش حاضر با هدف ارزیابی توانایی مدل درختی M5 در تخمین تبخیر-تعرق مرجع با استفاده از روش‌های پنمن-ماتنیث و هارگریوز-سامانی بود و موثرترین متغیرهای موثر در تخمین تبخیر-تعرق مرجع با استفاده از آنالیز حساسیت در دوایستگاه شیراز و کرمانشاه را تعیین نمود

مواد و روش‌ها

در این پژوهش از داده روزانه ایستگاه هواشناسی شیراز (به ارتفاع ۱۴۸۸ متر و عرض جغرافیایی $29^{\circ}36'N$ و طول جغرافیایی $52^{\circ}32'E$) و ایستگاه هواشناسی کرمانشاه (به ارتفاع ۱۳۱۸.۶ متر و عرض جغرافیایی $34^{\circ}21'N$ و طول جغرافیایی $49^{\circ}09'E$) طی دوره ۲۰۰۲-۱۹۹۵ استفاده شده شامل متوسط دمای هوا ($^{\circ}C$)، دمای نقطه شبنم ($^{\circ}C$)، متوسط رطوبت نسبی (%)، ساعات آفتابی (ساعت)، متوسط سرعت باد (متربرثانیه)، بارش (میلی-متر)، فشار بخار واقعی (میلی‌بار) و تبخیر-تعرق (میلی‌متر بر روز) محاسبه شده از رابطه‌ی پنمن-ماتنیث و هارگریوز-سامانی بود. از ۷۰٪ داده برای آموزش مدل با استفاده از مدل درختی M5 و از ۳۰٪ سری داده برای صحت‌سنجی مدل استفاده گردید. به منظور اجرای مدل از

^۱ WEKA

مورد نظر می‌رسند تغییرات ناچیزی داشته باشد و یا تعداد کمی از نمونه‌ها باقی بمانند، متوقف می‌شود. کاهش انحراف معیار با استفاده از رابطه زیر محاسبه می‌شود.

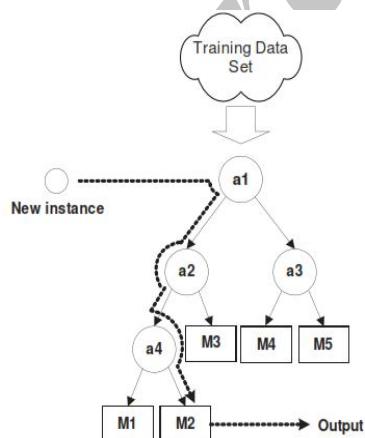
$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (3)$$

که T بیانگر نمونه‌هایی است که به گره می‌رسد؛ T_i

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad \text{بیانگر مجموعه-هایی هستند که}$$

از تقسیم کردن گره براساس صفت انتخابی به دست آمده-اند؛ sd بیانگر انحراف معیار است.

پس از ایجاد درخت، برای هر گره داخلی، یک مدل رگرسیون خطی چندگانه ساخته می‌شود. این مدل براساس داده‌های وابسته به آن گره و تمام صفاتی که در زیر درخت با ریشه آن گره مورد استفاده قرار گرفته‌اند، ایجاد می‌شود. در مرحله بعد، مدل‌های رگرسیون خطی، با کنارگذاشتن صفاتی که حذف آنها باعث کاهش میانگین خطا می‌شود، ساده‌سازی می‌شوند. بعد از این ساده‌سازی، هر زیر درخت برای انجام عملیات هرس کردن مورد بررسی قرار می‌گیرد. اگر خطای تخمین زده شده برای مدل خطی در ریشه زیر درخت، کوچک‌تر یا مساوی میانگین خطای درخت باشد، زیردرخت هرس می‌شود. شکل ۱ نحوه‌ی انشعاب مدل را بصورت شماتیک نشان می‌دهد



شکل ۱- ترکیب سلسله مراتبی مدل‌ها، a_i ها گره‌های انشعاب و M_j ها مدل‌ها

خاک (G) در دوره زمانی محاسبه ۲۴ ساعت صفر در نظر گرفته شد (آلن و همکاران، ۱۹۹۸). T_n, T_x, RH برای محاسبه e_s, e_a و استفاده شده‌اند. تابش روزانه خورشید (R_s) با استفاده از فرمول آنگستروم که تابش خورشید را به تابش برون‌زمینی و نسبت مدت روشنایی خورشید مرتبط می‌کند محاسبه گردید. روش هارگریوز-سامانی از طریق معادله زیر تعریف شد (هارگریوز، ۱۹۹۴):

$$ET_0 = 0.0023 R_a (T_a + 17.8) (T_x - T_n)^{0.5} \quad (2)$$

که ET_0 تبخیر تعرق گیاه مرجع محاسبه شده بوسیله معادله هارگریوز ($mm.day^{-1}$), R_a , تابش برون‌زمینی معادل آب ($mm.day^{-1}$) محاسبه شده طبق آنچه توسط آلن و همکاران (۱۹۹۸) محاسبه شده و T_a, T_n, T_x دمای بیشینه، کمینه و متوسط دمای روزانه است.

مدل درختی M5

الگوریتم M5 رایج‌ترین الگوریتم از خانواده درخت‌های تصمیم‌گیری می‌باشد. مدل درختی M5 یک الگوریتم پیش‌بینی عددی می‌باشد (ژانگ و تسای، ۲۰۰۷). مدل‌های درختی همانند درخت‌های رگرسیونی برای داده‌های زیاد موثر می‌باشند. به هر حال مدل‌های درختی معمولاً بسیار کوچک‌تر و دقیق‌تر از درخت‌های رگرسیونی می‌باشند (کوئینلن، ۱۹۹۲). در ابتدا الگوریتم مدل درختی M5 با تقسیم کردن فضای نمونه، بصورت بازگشتی یک درخت رگرسیونی می‌سازد. این تقسیم‌سازی برای کمینه کردن تغییرات زیرمجموعه‌های درونی مقادیر از ریشه تا گره و از مسیر شاخه‌ها صورت می‌گیرد. تغییرات بوسیله‌ی انحراف معیار مقادیری که از ریشه به هر گره رسیده‌اند، با محاسبه‌ی کاهش خطای مورد انتظار که از آزمایش هرصفت در آن گره حاصل می‌شود، اندازه‌گیری می‌شود. صفتی که میزان کاهش خطای مورد انتظار را بیشینه کند، انتخاب می‌شود. فرآیند جداسازی در صورتی که مقادیر خروجی تمامی نمونه‌هایی که به گره

خطای مورد انتظار برای داده آزمایشی در یک گره همانطور که قبلاً شرح داده شد با استفاده از مدل خطی برای پیش‌بینی محاسبه می‌شود. بدلیل فاکتور جبران $(n-v)/(n+v)$ ، مدل خطی می‌تواند با حذف عبارات به منظور حداقل نمودن خطای تخمینی ساده شود. حذف نمودن یک عبارت، فاکتور ضرب را که برای جبران نمودن افزایش ناگزیر در متوسط خطا در نمونه آموزشی کافی می‌باشد، کاهش می‌دهد. عبارات یکی یکی حذف می‌شوند تا تخمین خطا کاهش یابد.

در نهایت وقتی یک مدل خطی برای هر گره داخلی یافت شد، درخت از برگ‌ها به سمت عقب هرس می‌شود تا خطای تخمینی مورد انتظار کاهش یابد. خطای مورد انتظار برای مدل خطی در آن گره با خطای مورد انتظار از درخت فرعی زیرین مقایسه می‌شود. برای محاسبه مورد اخیر، خطای هر شاخه در یک مقدار تنها و کلی برای گره‌ای که با وزن‌دهی شاخه با نسبتی که نمونه‌های آموزشی که به آن می‌رسند ترکیب می‌شود و ترکیب تخمین خطاها بصورت خطی از این وزن‌ها استفاده می‌کند (ویتن و فرنک، ۲۰۰۵).

نتایج و بحث

برای تعیین تناسب مدل درختی M5 در پیش-بینی تبخیر-تعرق مرجع از سری داده‌ی شیراز و کرمانشاه استفاده شد. مقادیر تبخیر-تعرق مرجع برآورد شده از مدل درختی M5 با مقادیر تبخیر-تعرق محاسبه شده‌ی روش‌های پنمن-مانتیت و هارگریوز-سامانی مقایسه شدند. ضریب همبستگی، ریشه متوسط مربع خطا (RMSE) و مقادیر R، برای مقایسه‌ی عملکرد مدل درختی M5 با روش‌های پنمن-مانتیت و هارگریوز-سامانی استفاده شد که در جدول ۱ نشان داده شد. همانطور که در جدول مشاهده شد در ایستگاه شیراز مدل درختی M5 با مقدار ضریب همبستگی ۰/۹۸۸ و RMSE برابر ۰/۳۴۶ مقدار تبخیر-تعرق مرجع (روش پنمن-مانتیت) را برآورد نمود و با مقدار ضریب همبستگی ۰/۹۱۵ و مقدار RMSE برابر ۰/۸۴۴ مقدار تبخیر-تعرق مرجع (روش هارگریوز-

بعد از عملیات هرس کردن، ناپیوستگی زیادی بین مدل‌های خطی مجاور در برگ‌های درخت هرس شده، ایجاد می‌شود. در فرآیند هموارسازی در M5، مدل نهایی در یک برگ از ترکیب کردن مدل بدست آمده در آن برگ با مدل‌های موجود در مسیر ریشه تا برگ مربوطه بدست می‌آید. آزمایش‌های انجام شده توسط Wang و Witten نشان داده است که هموارسازی به میزان زیادی دقت پیش‌گویی‌ها را بهبود می‌بخشد (بنکدار و شهیدی، ۱۳۸۸).

همانطور که قبلاً بیان شد، یک مدل خطی برای هر گره داخلی، برای فرآیند هموارسازی مورد نیاز است. قبل از هرس نمودن یک مدل برای هر گره درخت هرس نشده محاسبه می‌شود. مدل شکل کلی زیر را دارد:

$$W_0 + W_1 a_1 + W_2 a_2 + \dots + W_k a_k \quad (4)$$

که a_1, a_2, \dots, a_k مقادیر صفات می‌باشند. وزن‌های W_1, W_2, \dots, W_k با استفاده از رگرسیون استاندارد محاسبه شده‌اند. اما فقط صفاتی که در درخت فرعی زیر این گره آزمون شده‌اند در رگرسیون استفاده می‌شوند زیرا سایر صفاتی که بر پیش‌بینی اثر می‌گذارند در تست‌هایی که به گره می‌رسند در نظر گرفته شده‌اند. در نظر داشته باشید که بطور ضمنی فرض شده است که صفات عددی هستند.

فرآیند هرس کردن از یک تخمین در هر گره، به عنوان خطای مورد انتظار برای داده آزمایشی استفاده می‌کند. ابتدا اختلاف مطلق بین مقدار پیش‌بینی شده و مقدار واقعی در هر نمونه آزمایشی که به گره می‌رسد، میانگین-گیری می‌شود. چون درخت دقیقاً برای این سری داده ساخته شده است، این میانگین خطای مورد انتظار برای مواردی که در نظر گرفته نشده‌اند را کم‌برآوردی می‌کند. برای جبران آن فاکتور $(n-v)/(n+v)$ که در آن n ، تعداد نمونه‌های آموزشی است که به گره می‌رسد و v ، تعداد پارامترهایی است که در مدل خطی مقدار کلاس در آن گره را می‌دهد، ضرب می‌شود.

در ایستگاه شیراز، مدل درختی M5 روش پنمن-مانتیت را با عملکرد بالاتری (R بیشتر و $RMSE$ کمتر) برآورد نمود ولی روش هارگریوز-سامانی در ایستگاه کرمانشاه عملکرد بالاتری نسبت به ایستگاه شیراز داشت.

سامانی) را محاسبه نمود. در ایستگاه کرمانشاه تبخیر-تعرق مرجع (روش پنمن-مانتیت) با ضریب همبستگی ۰/۹۸۶ و $RMSE$ برابر ۰/۳۶۲ و تبخیر-تعرق مرجع (روش هارگریوز-سامانی) با ضریب همبستگی ۰/۹۸۹ و مقدار $RMSE$ برابر ۰/۷۷۵ برآورد شد. به عبارتی

جدول ۱- مقایسه ضریب همبستگی و ریشه متوسط مربع خطای مدل درختی M5 با روش های پنمن-مانتیت و هارگریوز-سامانی در دو ایستگاه شیراز و کرمانشاه

ایستگاه	روش	R	$RMSE$
شیراز	پنمن-مانتیت	۰/۹۸۸	۰/۳۴۶
شیراز	هارگریوز-سامانی	۰/۹۱۵	۰/۸۴۴
کرمانشاه	پنمن-مانتیت	۰/۹۸۶	۰/۳۶۲
کرمانشاه	هارگریوز-سامانی	۰/۹۸۹	۰/۷۷۵

گمشده باشد، توسط روش‌های مختلف مانند استفاده از الگوریتم‌هایی که امکان جایگزینی مقدار گمشده را فراهم می‌کنند، آن مقدار یا مقادیر پر کند در صورتی که در روش‌های مستقیم، در صورتی که داده گمشده باشد امکان دسترسی به آن وجود نخواهد داشت و یا به ناچار توسط روش‌های میان‌یابی خطی جایگزین می‌شود.

$$LM \ 1: ET_o = 0.0599 * n + 0.0805 * T_{mean} - 0.0278 * Dew + 0.0891 * wind + 0.2791$$

$$LM \ 2: ET_o = 0.0343 * n + 0.0519 * T_{mean} - 0.0007 * Dew + 0.206 * wind + 0.3324$$

$$LM \ 3: ET_o = 0.0376 * n + 0.0818 * T_{mean} - 0.0013 * Dew + 0.1848 * wind - 6.6836$$

آنالیز حساسیت

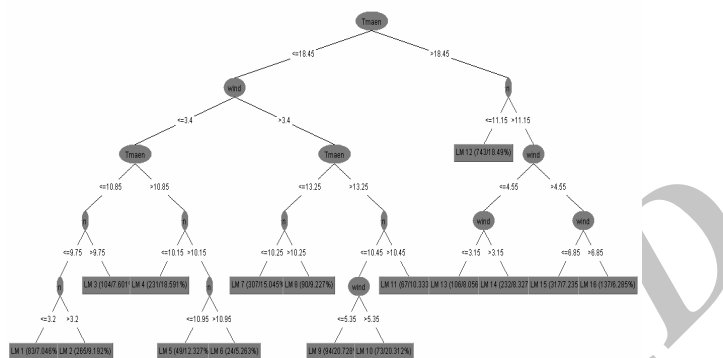
به منظور تعیین مهم‌ترین متغیرهای موثر در مدل‌سازی تبخیر-تعرق مرجع با بکارگیری مدل درختی M5 و روش پنمن-مانتیت، ابتدا طبق جدول ۲ برای ایستگاه شیراز، از تمامی متغیرهای مندرج در ردیف اول استفاده گردید و مقادیر ضریب همبستگی (R) و ریشه مربع متوسط خطا ($RMSE$) محاسبه شد. پس از آن داده بارش حذف شد و دوباره مدل اجرا گردید که تغییری در ضریب همبستگی مشاهده نشد اما مقدار $RMSE$ کاهش یافت که بطور کلی از متغیرهای ورودی حذف گردید. به

یکی از مزیت‌های مدل درختی M5 دسترسی آن به ترکیب چندین رابطه خطی ساده است که بتواند برای پیش‌بینی تبخیر-تعرق مرجع استفاده کرد. به عبارت دیگر می‌توان پی برد در شرایط مختلف دما، رطوبت، سرعت باد... چه معادله‌ای حاکم است و مهم‌ترین متغیرهایی که باعث انشعابات اولیه مدل می‌شوند کدام‌ها هستند. شکل ۲ مربوط به الگوریتم برآورد تبخیر-تعرق مرجع به روش پنمن-مانتیت در ایستگاه شیراز است که از اجرای مدل درختی M5 در نرم‌افزار وکا حاصل شد، نتیجه می‌شود اولین پارامتر مهم در ساخت مدل درختی M5، متوسط دمای روزانه است که در ابتدا درخت را به دو نیمه‌ی $T < 18/45$ و $T \geq 18/45$ تقسیم نمود. پس از آن طبق معیار انشعاب در $T \geq 18/45$ ، ساعات روشنایی از اهمیت بیشتری برخوردار است و در $T < 18/45$ ، متوسط سرعت باد اهمیت بیشتری دارد و انشعابات به همین صورت ادامه یافت تا در نهایت ۱۶ رابطه که هر کدام برای شرایط خاصی صدق می‌کنند بدست آمد. سه مورد از معادلات رگرسیونی حاصل شده از اجرای مدل در نرم‌افزار وکا آورده شده است.

مزیت دیگر الگوریتم‌های داده‌کاوی از جمله مدل درختی M5 آن است که امکان می‌دهد در صورتی که داده مربوط به ساعت، روز و یا ماه و یا هر مقیاس دیگری

آفتابی، سرعت باد، متوسط رطوبت نسبی و متوسط دمای نقطه شبنم بود که با وجود اینکه رطوبت نسبی نیز در آن لحاظ شده اما باز هم ضریب همبستگی کمتر و خطای بیشتری نسبت به ایستگاه شیراز داشت.

همین ترتیب تأثیر هر کدام از متغیرهای ورودی بر تبخیر-تعرق مرجع آزمون شد و متغیرهای دما، ساعات آفتابی، دمای نقطه شبنم و سرعت باد به عنوان موثرترین متغیرها معرفی شدند. در ایستگاه کرمانشاه نیز طبق جدول ۳، بهترین ترکیب استفاده از متوسط دمای هوا، ساعات



شکل ۲- الگوریتم مدل درختی M5 ایستگاه شیراز

شبنم و فشاربخار واقعی از موثرترین متغیرها بودند. ایستگاه کرمانشاه عملکرد بالاتری نسبت به ایستگاه شیراز نشان داد اما همانطور که ملاحظه می شود تعداد متغیرهای مؤثر نهایی آن نیز در برآورد تبخیرتعرق بیشتر بود.

در روش هارگریوز-سامانی نیز در ایستگاه شیراز متغیرهای متوسط دمای روزانه، ساعات آفتابی، متوسط رطوبت نسبی و فشاربخار واقعی و در ایستگاه کرمانشاه متغیرهای متوسط دمای روزانه، ساعات آفتابی، سرعت باد، متوسط رطوبت نسبی، متوسط دمای نقطه

جدول ۲- تحلیل حساسیت معادله پنمن - مانیتث ایستگاه شیراز

ترکیب متغیرهای ورودی	R	RMSE	MAE
T,n,RH,Dew,e,w,R	۰/۹۸۸	۰/۳۴۶	۰/۲۵۳
T,n,RH,Dew,e,w	۰/۹۸۸	۰/۳۴۰	۰/۲۴۹
T,n,RH,Dew,e	۰/۹۳۹	۰/۷۵۸	۰/۵۸۹
T,n,RH,Dew,w	۰/۹۸۸	۰/۳۴۴	۰/۲۵۲
T,n,RH,e,w	۰/۹۸۶	۰/۳۶۸	۰/۲۵۳
T,n,Dew,e,w	۰/۹۸۸	۰/۳۴۶	۰/۲۴۹
* T,n,dew,w	۰/۹۸۸	۰/۳۴۶	۰/۲۴۹
T,n,w	۰/۹۸۷	۰/۳۵۱	۰/۲۵۹
T,n	۰/۹۴۰	۰/۷۵۲	۰/۵۸۹
T,w	۰/۹۶۶	۰/۵۷۴	۰/۴۵۹
n,w	۰/۸۶۳	۱/۱۱۹	۰/۸۳۸

T (°C): متوسط دمای هوا، n (ساعت): ساعات روشنایی، RH (%): متوسط رطوبت نسبی،

Dew (°C): متوسط دمای نقطه شبنم، e (میلی بار): فشاربخار واقعی، w (متر بر ثانیه):

متوسط سرعت باد، R (میلی متر): بارش. * بهترین ترکیب متغیرها

جدول ۳- تحلیل حساسیت معادله پنمن - ماتیت ایستگاه کرمانشاه

ترکیب متغیرهای ورودی	R	RMSE	MAE
T,n,w,RH,dew,e,R	۰/۹۸۶	۰/۳۶۳	۰/۲۷۴
T,n,w,RH,dew,e	۰/۹۸۶	۰/۳۶۱	۰/۲۷۴
* T,n,w,RH,dew	۰/۹۸۶	۰/۳۶۲	۰/۲۷۳
T,n,w,RH,e	۰/۹۸۶	۰/۳۶۳	۰/۲۷۳
T,n,w,dew,e	۰/۹۸۵	۰/۳۷۱	۰/۲۸۲
T,n,RH,dew,e	۰/۹۴۳	۰/۷۲۹	۰/۵۶۰
T,n,w,RH	۰/۹۸۶	۰/۳۶۸	۰/۲۷۷
T,n,w,dew	۰/۹۸۵	۰/۳۷۵	۰/۲۸۳
T,n,w,e	۰/۹۸۵	۰/۳۷۵	۰/۲۸۴
T,n,w	۰/۹۸۳	۰/۳۹۸	۰/۳۰۳
T,n	۰/۹۴۰	۰/۷۴۲	۰/۵۷۷

جدول ۴- تحلیل حساسیت معادله هارگریوز - سامانی ایستگاه شیراز

ترکیب متغیرهای ورودی	R	RMSE	MAE
T,n,RH,Dew,e,w,R	۰/۹۰۸	۰/۸۷۴	۰/۶۰۹
T,n,RH,Dew,e,w	۰/۹۰۹	۰/۸۷۳	۰/۶۰۹
T,n,RH,Dew,e	۰/۹۱۵	۰/۸۴۶	۰/۶
T,n,RH,Dew,w	۰/۹۱۰	۰/۸۶۶	۰/۶۰۵
T,n,RH,e,w	۰/۹۱۳	۰/۸۵۱	۰/۶۰۶
T,n,Dew,e,w	۰/۹۱۴	۰/۸۴۸	۰/۵۹۸
T,RH,Dew,e,w	۰/۸۹۴	۰/۹۳۴	۰/۶۵۱
T,n,Dew,e	۰/۹۱۴	۰/۸۵۱	۰/۶
* T,n,RH,e	۰/۹۱۵	۰/۸۴۴	۰/۶۰۱
T,n,e	۰/۹۱۴	۰/۸۵۱	۰/۹۱۴
T,n,RH	۰/۹۱۳	۰/۸۵۳	۰/۶۰۱

جدول ۵- تحلیل حساسیت معادله هارگریوز - سامانی ایستگاه کرمانشاه

ترکیب متغیرهای ورودی	R	RMSE	MAE
T,n,w,RH,dew,e,R	۰/۹۸۹	۰/۷۷۴	۰/۵۵۵
* T,n,w,RH,dew,e	۰/۹۸۹	۰/۷۷۴	۰/۵۵۶
T,n,w,RH,dew	۰/۹۸۹	۰/۷۷۸	۰/۵۵۸
T,n,w,RH,e	۰/۹۸۹	۰/۷۸۳	۰/۵۵۹
T,n,w,dew,e	۰/۹۸۹	۰/۷۹۱	۰/۵۶۸
T,n,RH,dew,e	۰/۹۸۹	۰/۸۰۵	۰/۵۷۸
T,n,w,RH	۰/۹۸۹	۰/۷۸۹	۰/۵۶۳
T,n,w,dew	۰/۹۸۹	۰/۷۹۳	۰/۵۷۱
T,n,w,e	۰/۹۸۹	۰/۷۹۳	۰/۵۷۱
T,n,w	۰/۹۸۹	۰/۸۱۲	۰/۵۸۶
T,n	۰/۹۸۸	۰/۸۳۲	۰/۶۰۴

نتیجه گیری

عملکرد مدل در تخمین تبخیر- تعرق مرجع در روش هارگریوز- سامانی بالاتر بود اما از تعداد متغیرهای بیشتری نیز استفاده شد. هم‌چنین مدل درختی نشان داد متوسط دمای روزانه مهم‌ترین متغیر در تبخیر-تعرق در روش پنمن-مانتیت است. در ایستگاه شیراز در روش پنمن-مانتیت، آنالیز حساسیت نشان داد که مهم‌ترین متغیرهای موثر بر آن در روش پنمن-مانتیت، ساعات آفتابی، متوسط دمای روزانه، دمای نقطه شبنم و سرعت باد خواهد بود اما در روش هارگریوز-سامانی مهم‌ترین متغیرهای موثر عبارتند از ساعات آفتابی، متوسط دمای هوا، متوسط رطوبت نسبی و فشار بخار واقعی. در ایستگاه کرمانشاه در روش پنمن-مانتیت موثرترین متغیرها عبارتند از متوسط دمای روزانه، ساعات آفتابی، سرعت باد، متوسط رطوبت نسبی و متوسط دمای نقطه شبنم و در روش هارگریوز- سامانی عبارتند از همان متغیرهای بکار گرفته شده در روش پنمن-مانتیت بعلاوه فشاربخار.

نتایج این پژوهش نشان داد که با استفاده از تکنیک‌های داده‌کاوی از جمله مدل‌های درختی می‌توان در برآورد تبخیر- تعرق مرجع استفاده نمود. هم‌چنین می‌توان بجای استفاده از متغیرهای هواشناسی زیاد بکارگرفته در روابط معرفی شده توسط محققین مختلف، از موثرترین آنها استفاده کرد و به همان دقت دست یافت. در صورتی که ایستگاه مورد نظر مجهز به ایستگاه خودکار هواشناسی باشد، به جای محاسبه‌ی تبخیر- تعرق بصورت انجام شده در این مقاله، می‌توان از مقادیر ثبت شده توسط دستگاه استفاده نمود. در این مقاله توانایی مدل درختی M5، در تخمین تبخیر- تعرق مرجع در مقایسه با روش-های پنمن-مانتیت و هارگریوز-سامانی ارزیابی گردید. همانطور که نتایج نشان داد، در ایستگاه شیراز مدل درختی M5، توانست تبخیر- تعرق مرجع به هر دو روش را بخوبی برآورد نماید اما توانایی تخمین آن به روش هارگریوز-سامانی کمتر است. در ایستگاه کرمانشاه

فهرست منابع:

- Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. ۱۹۹۸. Crop evapotranspiration, Irrigation and Drainage. Paper No. ۵۶. FAO, Rome, Italy.
- Bonakdar, L. and Etemad-Shahidi, A. (۲۰۱۱), Predicting wave run-up on rubble-mound structures using M^۵ model tree, Ocean engineering, ۳۸, ۱۱۱-۱۱۸
- Amatya, D.M., Skaggs, R.W. and Gregory, J.D. ۱۹۹۵. Comparison of methods for estimating REF-ET. Irrigation and Drainage Engineering. ۱۲۱(۶): ۴۲۷-۴۳۵.
- Berengena, J. and Gavilan, P. ۲۰۰۵. Reference evapotranspiration estimation in a highly advective semiarid environment. Irrigation and Drainage Engineering. ASCE, ۱۳۱(۲): ۱۶۳-۱۶۷.
- Braha, D. and Shmilovici, A. ۲۰۰۲. Data mining for improving a cleaning process in the semiconductor industry. IEEE Transactions on Semiconductor Manufacturing, ۱۵: ۹۱-۱۰۱.
- Choisnel, E., Gavilan, P. ۱۹۹۲. Une approche uniformisee du calcul l'evapotranspiration potentielle pour l'ensemble des pays de la communaut'es europ'eene. Publication EUR ۱۴۲۲۳, Office des publications officielles des communaut'e europ'eene: Luxemburg.
- Droogers, P. and Allen, R.G. ۲۰۰۲. Estimating reference evapotranspiration under inaccurate data conditions. Irrigation and Drainage Systems. ۱۶: ۳۳-۴۵.
- Fayyad, U.M. and Uthurusamy, R. ۲۰۰۲. Evolving data mining into solutions for insights. Commun. ACM. ۴۵: ۲۸-۳۱.
- Goodwin, L., VanDyne, M., Lin, S., and Talbert, S. ۲۰۰۳. Data mining issues and opportunities for building nursing knowledge. Biomed. Inform. ۳۶: ۳۷۹-۳۸۸.

۱۰. Hargreaves, G.L. and Samani, Z.A. ۱۹۸۲. Estimating potential evapotranspiration. *Irrigation and Drainage Engineering*. ۱۰۸(۳): ۲۲۵-۲۳۰.
۱۱. Hargreaves, G.H. and Allen, R.G. ۲۰۰۳. History and evaluation of Hargreaves evapotranspiration equation. *Irrigation and Drainage Engineering*. ۱۲۹(۱): ۵۳-۶۳.
۱۲. Hargreaves, G. H. ۱۹۹۴. Defining and using reference evapotranspiration. *Irrigation and Drainage Engineering*. ۱۲۰(۶): ۱۱۳۲-۱۱۳۹.
۱۳. Jensen, M.E., Burman, R.D. and Allen, R.G. ۱۹۹۰. Evapotranspiration and irrigation water requirements. ASCE Manual and Rep. on Engrg. Pract. No. ۷۰, ASCE, NY
۱۴. Katul, G.G., Cuenca, R.H., Grebet, P., Wright, J.L. and Pruitt, W.O. ۱۹۹۲. Analysis of evaporative flux data for various climates. *Irrigation and Drainage Engineering*. ۱۱۸(۴): ۶۰۱-۶۱۸.
۱۵. Li, S.T. and Shue, L.Y. ۲۰۰۴. Data mining to aid policy making in air pollution management. *Expert Systems with Applications*. ۲۷: ۳۳۱-۳۴۰.
۱۶. Meyer, S.J., Hubbard, K.G. and Wilhite, D.A. ۱۹۸۹. Estimating potential evapotranspiration: the effect of random and systematic errors. *Agricultural and Forest Meteorology*. ۴۶: ۲۸۵-۲۹۶.
۱۷. Pal, M. and Deswal, S. ۲۰۰۹. M^۵ model tree based modeling of reference evapotranspiration. *Hydrological Processes*. ۲۳: ۱۴۳۷-۱۴۴۳.
۱۸. Pal, M. ۲۰۰۶. M^۵ model tree for land cover classification. *Remote Sensing*. ۲۷(۴): ۸۲۵-۸۳۱.
۱۹. Quinlan, J.R. ۱۹۹۲. Learning with continuous classes. *Proceedings of Australian Joint Conference on Artificial Intelligence*. World Scientific Press: Singapore; ۳۴۳-۳۴۸.
۲۰. Sharma, M.L. ۱۹۸۵. Estimating evapotranspiration. In: Hillel D (ed). *Advances in Irrigation*, vol ۳. Acain, London.
۲۱. Smith, M., Allen, R. and Pereira, L. ۱۹۹۶. Revised FAO methodology for crop water requirements. In: *Proceedings of the International Conference on evapotranspiration and Irrigation Scheduling*, San Antonio, TX, USA, pp. ۱۳۳-۱۴۰.
۲۲. Terzi, Ö. ۲۰۰۷. Data mining approach for estimation evaporation from free water surface. *Applied Sciences*. ۷(۴): ۵۹۳-۵۹۶.
۲۳. Terzi, Ö. And Keskin, M. E. ۲۰۰۵. Evaporation estimation using gene expression programming. *Applied Sciences*. ۵: ۵۰۸-۵۱۲.
۲۴. Ventura, F., Spano, D., Duce, P., Snyder, R.L. ۱۹۹۹. An evaluation of common evapotranspiration equations. *Irrigation Sciences*. ۱۸: ۱۶۳-۱۷۰.
۲۵. Witten, I.H. and Frank, E. ۲۰۰۵. *Data Mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann, San Francisco.