



## Investigating the effectiveness of data mining methods in predicting daily reference evapotranspiration (Case study: Coastal strip stations in southern Iran)

Halimeh Piri 

Associate Professor, Department of Water Engineering, Faculty of Water and Soil, University of Zabol  
Email: [h\\_piri2880@uoz.ac.ir](mailto:h_piri2880@uoz.ac.ir)

### ARTICLE INFO

**Article type:**  
Research Article

**Article History:**  
Received March 25, 2024  
Revised May 07, 2024  
Accepted June 05, 2024

**Keywords:**  
*decision tree,*  
*gaussian process*  
*regression,*  
*random forest,*  
*support vector machine,*  
*Principal Component*  
*Analysis.*

### ABSTRACT

Reference evapotranspiration is one of the important hydrological parameters in agricultural research and water and soil protection. For this reason, its estimation and modeling are of special importance. Non-linear relationships, inherent uncertainty, and the need for a lot of climate information in estimating evapotranspiration have made researchers use data-mining methods to estimate evapotranspiration in recent decades. The purpose of this research is to investigate the efficiency of data mining methods included in support vector machine, decision tree, random forest and Gaussian process regression in forecasting the daily reference evapotranspiration of coastal strip stations in the south of the country (Chabahar, Bandar Abbas, Bushehr and Abadan). To do the work, daily reference evapotranspiration was calculated using 20-year climatic data (2001-2021) using the FAO-Penman-Monteith method. Then, using these data as output data, 6 combined scenarios were evaluated based on the correlation between meteorological variables and reference evapotranspiration using data mining methods (Support vector machine, Decision tree, Random forest and Gaussian regression process). The results of the investigations showed that all four data mining methods were able to estimate the reference evapotranspiration values in the studied areas. In all four stations, the Gaussian process regression method with the highest R<sup>2</sup> value and the lowest RMSE and MAE values had a better estimate of the reference evapotranspiration values, and random forest, decision tree, and support vector machine methods were in the next ranks respectively. Among the examined patterns, in Chabahar pattern 6 with the combination of minimum temperature, maximum temperature, average temperature, relative humidity, solar radiation and wind speed variables, in Bandar Abbas and Bushehr pattern 4 with the combination of minimum temperature, maximum temperature, average temperature and relative humidity and in Abadan pattern 3 with the combination of minimum temperature, maximum temperature and average temperature parameters had the best estimate. Considering the high accuracy of the Gaussian process regression model in estimating reference evapotranspiration, this method is recommended for estimating reference evapotranspiration in coastal stations of southern Iran.

**Cite this article:** Piri, H. (2024). Investigating the effectiveness of data mining methods in predicting daily reference evapotranspiration (Case study: Coastal strip stations in southern Iran). *ECO HYDROLOGY*. 11 (2), 271-286. Doi: [doi.org/10.22059/IJE.2024.375755.1816](https://doi.org/10.22059/IJE.2024.375755.1816)





## بررسی کارایی روش‌های داده‌کاوی در پیش‌بینی تبخیر-تعرق مرجع روزانه (مطالعه موردی: ایستگاه‌های نوار ساحلی جنوب ایران)

حلیمه پیری

دانشیار، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران. رایانامه: [h\\_piri2880@uoz.ac.ir](mailto:h_piri2880@uoz.ac.ir)

چکیده	اطلاعات مقاله
تبخیر-تعرق مرجع یکی از پارامترهای مهم هیدرولوژیکی در تحقیقات کشاورزی و حفاظت آب و خاک است. به همین دلیل برآورد و مدل‌سازی آن از اهمیتی ویژه برخوردار است. روابط غیرخطی، عدم قطعیت ذاتی و نیاز به اطلاعات اقلیمی فراوان در برآورد تبخیر-تعرق باعث شده است پژوهشگران در دهه‌های اخیر از روش‌های داده‌کاوی برای برآورد تبخیر-تعرق استفاده کنند. هدف از این تحقیق بررسی کارایی روش‌های داده‌کاوی ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و رگرسیون فرایند گاوسی در پیش‌بینی تبخیر-تعرق مرجع روزانه برخی از ایستگاه‌های نوار ساحلی جنوب کشور شامل چابهار، بندرعباس، بوشهر و آبادان است. برای انجام کار با استفاده از داده‌های اقلیمی ۲۰ ساله (۱۳۸۰-۱۴۰۰)، تبخیر-تعرق مرجع روزانه به روش فائو-پنمن-مانتیت محاسبه شد. سپس با استفاده از این داده‌ها به‌عنوان داده‌های خروجی، ۶ سناریو ترکیبی براساس همبستگی بین متغیرهای هواشناسی و تبخیر-تعرق مرجع از طریق روش‌های داده‌کاوی ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و فرایند رگرسیون گاوسی مورد ارزیابی قرار گرفتند. نتایج بررسی‌ها نشان داد هر چهار روش داده‌کاوی در مناطق مورد مطالعه به‌خوبی توانسته‌اند مقادیر تبخیر-تعرق مرجع را برآورد کنند. در هر چهار ایستگاه، روش رگرسیون فرایند گاوسی با داشتن بالاترین مقدار $R^2$ و کمترین مقادیر RMSE و MAE برآورد بهتری از مقادیر تبخیر-تعرق مرجع داشتند و روش‌های جنگل تصادفی، درخت تصمیم و ماشین بردار پشتیبان به ترتیب در رتبه‌های بعدی قرار گرفتند. از بین الگوهای مورد بررسی، در چابهار الگوی ۶ با ترکیب متغیرهای دمای حداقل، دمای حداکثر، دمای میانگین، رطوبت نسبی، تابش خورشیدی و سرعت باد، در بندرعباس و بوشهر الگوی ۴ با ترکیب متغیرهای دمای حداقل، دمای حداکثر، دمای میانگین و رطوبت نسبی و در آبادان الگوی ۳ با ترکیب متغیرهای دمای حداقل، دمای حداکثر و دمای میانگین بهترین برآورد را داشتند. باتوجه‌به دقت بالای مدل رگرسیون فرایند گاوسی در برآورد تبخیر-تعرق مرجع، این روش برای برآورد تبخیر-تعرق مرجع در ایستگاه‌های نوار ساحلی جنوب ایران توصیه می‌شود.	<p>نوع مقاله: پژوهشی</p> <p>تاریخ‌های مقاله: تاریخ دریافت: ۱۴۰۳/۰۱/۱۵ تاریخ بازنگری: ۱۴۰۳/۰۲/۲۱ تاریخ پذیرش: ۱۴۰۳/۰۳/۱۸</p> <p>کلیدواژه‌ها: جنگل تصادفی، درخت تصمیم، رگرسیون فرایند گاوسی، ماشین بردار پشتیبان، مؤلفه‌های اصلی.</p>

استناد: پیری، حلیمه. (۱۴۰۳). بررسی کارایی روش‌های داده‌کاوی در پیش‌بینی تبخیر-تعرق مرجع روزانه (مطالعه موردی: ایستگاه‌های نوار ساحلی جنوب ایران). *اكو هيدرولوژي*، ۲۱(۲)، ۲۷۱-۲۸۶.

DOI: <http://doi.org/10.22059/IJE.2024.375755.1816>

ناشر: مؤسسه انتشارات دانشگاه تهران.

© حلیمه پیری.

DOI: <http://doi.org/10.22059/IJE.2024.375755.1816>



## ۱. مقدمه

تعیین آب مصرفی گیاهان باغی و زراعی در مناطق خشک و نیمه‌خشک، برنامه‌ریزی آبیاری، تقسیم آب، حبابه‌ها و تهیه و تنظیم اسناد ملی آبیاری همگی بر اساس تبخیر-تعرق است. همچنین در طراحی انواع سازه‌های انتقال آب، کنترل سیلاب در طرح‌های آبیاری و زهکشی و حتی مصارف آب شهری، نیاز مبرمی به آگاهی از مقدار صحیح تبخیر-تعرق است. برآورد تبخیر-تعرق مرجع در طرح‌های آبیاری و زهکشی برای تعیین میزان آب مصرفی طرح‌ها بسیار ضروری است. یکی از راهکارهای اساسی برای برآورد مقدار آب مورد نیاز محصولات کشاورزی، دانستن روابط بین وضعیت آب‌وهوایی و تبخیر-تعرق است [۱]. برآورد تبخیر-تعرق به روش‌های مختلف انجام می‌شود. یکی از روش‌ها ضرب تبخیر-تعرق مرجع در ضریب گیاهی است. بنابراین به دست آوردن تبخیر-تعرق مرجع ضروری است. مقدار تبخیر-تعرق گیاه مرجع را می‌توان به صورت مستقیم، به وسیله لایسیمتر و غیرمستقیم (روش‌های آیرودینامیک، توازن انرژی و ترکیبی) و یا از طریق روش‌های تجربی به دست آورد. به علت مشکلات احداث، نیاز به دقت بالا، انجام آزمایش‌های متعدد و هزینه نسبتاً زیاد، استفاده از روش‌های غیرمستقیم براساس اطلاعات هواشناسی ضرورت پیدا می‌کند. روش‌های تجربی براساس آنالیزهای آماری و مشاهدات ایستگاه‌های هواشناسی و معمولاً برای یک منطقه یا وضعیت آب‌وهوایی یا اقلیمی خاص ارائه می‌شوند [۲]. مدل‌های فیزیکی بیشتر هدفشان شبیه‌سازی اصول بنیادی مثل تعادل انرژی و جابجایی جرم است. برای مثال تابش خورشیدی، رطوبت نسبی و سرعت باد، تعدادی از داده‌های مهم برای محاسبه تبخیر-تعرق مرجع براساس مدل‌های فیزیکی هستند [۳].

روش فائو-پنمن-مانتیت، یکی از روش‌های ترکیبی برای تخمین تبخیر-تعرق گیاه مرجع است که در آن گیاه مرجع، گیاه فرضی (معمولاً چمن) با ارتفاع ۱۲ سانتی‌متر، ضریب بازتاب ۲۳ درصد و مقاومت روزنه ۷۰ ثانیه بر متر است. این روش از آن جهت حائز اهمیت است که بیش از آنکه براساس تجربه استوار باشد، مبنای فیزیکی داشته و اکثر عواملی که در تبخیر-تعرق مؤثرند، در آن دخالت داده شده است. کارشناسان سازمان فائو به منظور استاندارد کردن روش محاسبه تبخیر-تعرق، روش فائو-پنمن-مانتیت را به عنوان مبنای انتخاب کرده و براساس آن تبخیر-تعرق گیاه مرجع را برای گیاه فرضی محاسبه کرده‌اند [۴]. با توجه به تحقیقات صورت گرفته، روش فائو-پنمن-مانتیت یکی از دقیق‌ترین روش‌های تخمین تبخیر-تعرق است و می‌توان از آن برای اقلیم‌های مختلف استفاده کرد. با وجود این، این معادله نیاز به داده‌های اولیه فراوان (دمای هوا، تابش خورشید، سرعت باد و رطوبت نسبی) دارد [۵، ۶ و ۷].

تعداد زیاد داده‌ها در روش فائو-پنمن-مانتیت با وجود دقت قابل قبول آن، همواره پژوهشگران را به سوی استفاده از روش‌های ساده‌تر و در عین حال دقیق‌تر سوق داده است. بنابراین از مدل‌های گوناگونی برای برآورد تبخیر-تعرق استفاده می‌کنند و سعی در انتخاب مدلی با دقت بیشتر، خطای کمتر و در عین حال کمترین تعداد متغیر ورودی دارند. این مدل‌ها مشکلات و هزینه‌های روش‌های مستقیم را ندارند و می‌توان به منظور پیش‌بینی نیز از آن‌ها استفاده کرد.

نتایج تحقیقات نشان می‌دهد که مدل‌های خطی، به دلیل تغییرپذیری زیاد مؤلفه تبخیر-تعرق عملکرد مناسبی در برآورد آن نداشته و باید از مدل‌های غیرخطی استفاده کرد که خود نیازمند برآورد پارامترهای زیادی خواهد بود [۸]. در حال حاضر، استفاده از روش‌های داده‌کاوی مهم‌ترین روش در مدل‌سازی فرایندهای پیچیده و غیرخطی است و می‌تواند به منظور مدل‌سازی تبخیر-تعرق به کار گرفته شود [۹].

روش‌های داده‌کاوی برای مجموعه داده‌های بزرگ با متغیرهای زیاد ساخته شده‌اند. از جمله روش‌های داده‌کاوی می‌توان به مدل‌های ماشین بردار پشتیبان (SVM)،<sup>۱</sup> جنگل تصادفی (RF)،<sup>۲</sup> مدل درخت تصمیم (M5)<sup>۳</sup> و رگرسیون فرایند گاوسی (GPR)<sup>۴</sup> اشاره کرد. ماشین بردار پشتیبان یک طبقه‌بندی‌کننده دوتایی غیرآماری است که در سال‌های اخیر بسیار مورد توجه قرار گرفته است. رویکرد این مدل به این صورت است که در فاز آموزش، تلاش می‌شود که مرز تصمیم‌گیری به گونه‌ای انتخاب شود که حداقل فاصله آن با هر یک از دسته‌های مورد نظر ماکزیمم شود [۱۰]. از مزایای این روش آموزش نسبتاً ساده آن است، برخلاف شبکه‌های عصبی مصنوعی در ماکزیمم‌های محلی گیر نمی‌افتد، برای داده‌های با ابعاد بالا تقریباً خوب جواب می‌دهد [۱۱]. مدل جنگل تصادفی به عنوان روشی توسعه‌یافته برای بهبود و بهسازی روش‌های درخت طبقه‌بندی و درخت رگرسیون به کار گرفته می‌شود. این مدل دارای دقت و قابلیت انطباق و سازگاری تقویت شده است و در انجام محاسبات از سرعت قابل قبولی برخوردار است. همچنین در انجام فرایند مدل‌سازی با جنگل تصادفی نیازی به نرمال‌سازی داده‌ها

1. Support Vector Machine
2. Random Forest
3. Decision Tree
4. Gaussian Process Regression

وجود ندارد [۱۲]. مدل M5 یک مدل درختی برای پیش‌بینی صفات عددی پیوسته است که در آن، توابع رگرسیونی خطی در برگ‌ها ظاهر می‌شوند. نتایج مدل درختی برای فهم و شبیه‌سازی آسان است و خروجی مدل از دقت بالایی برخوردار است که می‌توان آن را با سایر مدل‌ها مقایسه کرد [۱۳]. رگرسیون فرایند گاوسی، تعمیم‌یافته توزیع گاوسی است. مدل‌های رگرسیون فرایند گاوسی قادر به درک توزیع پیش‌بینی متناظر با ورودی آزمون هستند. فرایندهای گاوسی همان قدر که ساده‌اند، فراگیر نیز هستند [۱۴].

تحقیقات زیادی در زمینه استفاده از روش‌های داده‌کاوی در پیش‌بینی تبخیر-تعرق مرجع در مناطق مختلف انجام شده است. در تحقیقی، تبخیر-تعرق مرجع روزانه تبریز با استفاده از روش‌های داده‌کاوی ماشین بردار پشتیبان و مدل درختی M5 برآورد شد. نتایج تحقیق نشان داد روش‌های داده‌کاوی نسبت به روش‌های تجربی، برآورد بهتری از تبخیر-تعرق مرجع داشتند [۱۵]. کارایی روش‌های ماشین بردار پشتیبان، جنگل تصادفی و مدل کیوبیست در برآورد تبخیر-تعرق مرجع بوشهر نسبت به مدل‌های تجربی مورد بررسی قرار گرفت. نتیجه مطالعه نشان داد هر سه مدل توانایی بالایی در برآورد تبخیر-تعرق مرجع نسبت به مدل‌های تجربی دارا بودند [۱۶]. همچنین با استفاده از پنج روش داده‌کاوی، تبخیر از ایستگاه هواشناسی تبریز برآورد شد. نتایج تحقیق نشان داد مدل ماشین بردار پشتیبان و مدل فرایند رگرسیون گاوسی، برآورد بهتری از مقدار تبخیر داشت [۱۷]. در برآورد تبخیر-تعرق مرجع بناب نیز از روش‌های شبکه عصبی مصنوعی و مدل درختی استفاده شد. نتیجه تحقیق نشان داد اگرچه روش شبکه عصبی مصنوعی با اختلاف خیلی کمی در مقایسه با روش درختی پیش‌بینی نسبتاً دقیق‌تری از میزان تبخیر-تعرق مرجع ارائه می‌کند، روش درختی، روابط ساده خطی قابل فهم‌تر و کاربردی‌تری برای پیش‌بینی تبخیر-تعرق ارائه می‌کند [۱۸]. از روش‌های درخت تصمیم، یادگیری عمیق و جنگل تصادفی برای برآورد تبخیر-تعرق مرجع روزانه در منطقه سیستان استفاده شد. نتایج تحقیق نشان داد مدل یادگیری عمیق به‌عنوان بهترین مدل در بین مدل‌های انتخابی است [۱۹]. در تحقیقی دیگر، سه روش داده‌محور درخت گرادیان تقویت‌شده<sup>۱</sup>، مدل خطی تعمیم‌یافته<sup>۲</sup> و پرسپترون چندلایه<sup>۳</sup> در برآورد مقدار تبخیر از تشت در سه اقلیم خشک (ایستگاه یزد و بافق)، نیمه‌خشک (ایستگاه بیرجند و سیاه‌بیشه) و مرطوب (ایستگاه ساری و فردوس) مورد بررسی قرار گرفت. نتایج تحقیق نشان داد در تمامی ایستگاه‌های مورد مطالعه، روش پرسپترون چندلایه دقیق‌ترین برآوردها را ارائه نمود و به‌عنوان روشی با دقت بالا پیشنهاد شد [۲۰]. بنابراین در یک جمع‌بندی می‌توان بیان نمود که اغلب پژوهش‌های انجام‌شده در زمینه برآورد تبخیر-تعرق یا با استفاده از معادلات تجربی است که دقت بالایی ندارند و یا با روابط غیرخطی پیشنهادی توسط پژوهشگران انجام شده است که نیازمند متغیرهای هواشناسی زیادی است تا تخمین قابل قبولی با خطای کمتر ارائه دهند؛ اما نتایج تحقیقات انجام‌شده به روش‌های داده‌کاوی نشان داد که این روش‌ها، روش‌های مناسبی برای تبخیر-تعرق مرجع هستند.

پژوهش‌های بسیاری در خصوص برآورد تبخیر-تعرق مرجع در شرایط کمبود داده‌های هواشناسی برای وضعیت آب‌وهوایی ایران انجام شده است [۲۱]؛ اما با وجود توسعه زیاد کشاورزی در مناطق ساحلی ایران و بالا بودن تبخیر-تعرق در سواحل جنوبی کشور، تحقیقات جامعی برای محاسبه مقدار تبخیر-تعرق در این مناطق صورت نگرفته است. طول خط ساحل کشور ایران حدود ۵۸۰۰ کیلومتر است که ۸۹۰ کیلومتر آن در شمال ایران و دارای اقلیم معتدل و مرطوب، شامل استان‌های گیلان، مازندران و گلستان است. ۴۹۰۰ کیلومتر خط ساحلی ایران نیز در جنوب کشور (استان‌های خوزستان، بوشهر، هرمزگان و سیستان و بلوچستان) واقع شده و دارای اقلیم گرم و مرطوب است. دریای خزر به اقیانوس راه ندارد و سطح آن نیز پایین‌تر از سطح دریاهای آزاد است. خط ساحلی جنوب ایران نیز از طریق دریای عمان به آب‌های آزاد راه دارد. فعالیت‌های کشاورزی زیادی در نوار ساحلی کشور صورت می‌گیرد [۲۲].

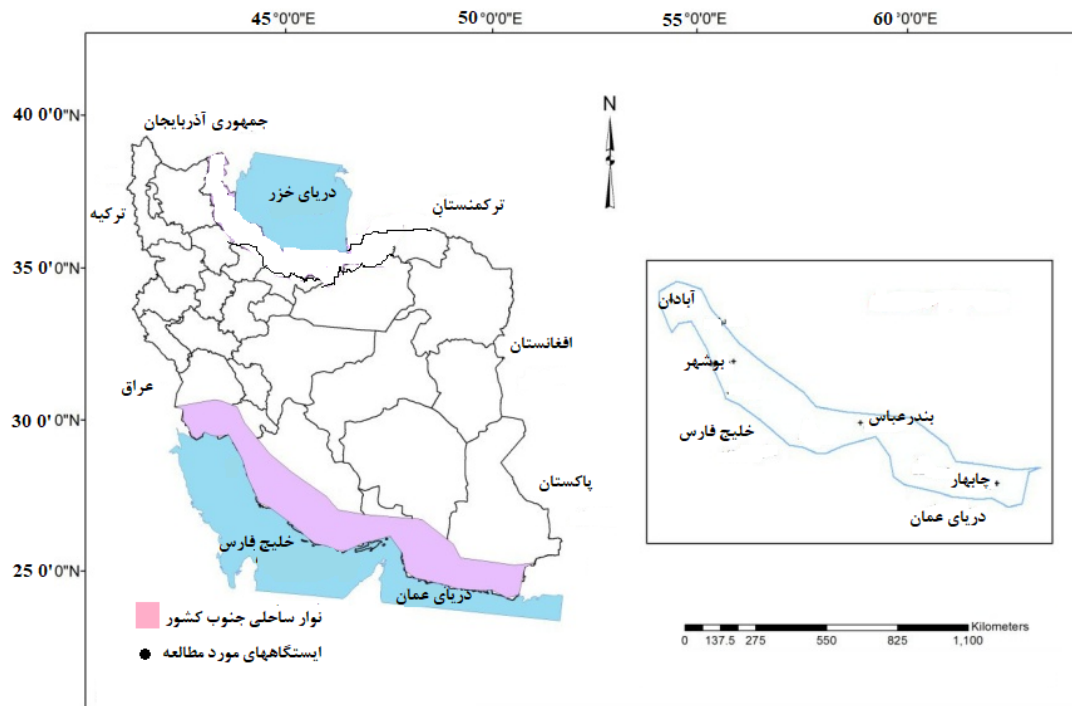
همان‌طور که گفته شد، دقت معادله فائو-پنمن-مانتیت برای محاسبه تبخیر-تعرق مناسب است؛ اما نیاز به داده‌ها و اطلاعات زیادی دارد که در بسیاری از موارد، به‌خصوص در کشورهای جهان سوم (مانند ایران) در دسترس نیست و در نتیجه از روش‌های جایگزین که از داده‌ها و اطلاعات کمتری استفاده می‌کنند، بیشتر استفاده می‌شود. بنابراین هدف از این تحقیق، بررسی چهار روش داده‌کاوی ماشین بردار پشتیبان، جنگل تصادفی، مدل درختی و مدل رگرسیون فرایند گاوسی با ترکیبات مختلف هواشناسی در برآورد مقدار تبخیر-تعرق مرجع روزانه بعضی ایستگاه‌های نوار ساحلی جنوب کشور شامل چابهار، بندرعباس، بوشهر و آبادان است.

1. Gradient boosted tree
2. Generalized linear model
3. Multi layer perceptron

## ۲. مواد و روش‌ها

## ۱.۲. مناطق مورد مطالعه

مناطق مورد مطالعه برخی ایستگاه‌های هواشناسی نوار ساحلی جنوب کشور شامل چابهار، بندرعباس، بوشهر و آبادان است. به‌منظور کاهش حجم مقاله و اطالعه مطالب، از هر چهار استان سواحل جنوبی، یک ایستگاه انتخاب شد. شکل ۱ موقعیت ایستگاه‌های مورد مطالعه را نشان می‌دهد.



شکل ۱: مناطق مورد مطالعه

برای انجام تحقیق ابتدا داده‌های هواشناسی ۲۰ ساله (۱۳۸۰-۱۴۰۰) از ایستگاه‌های مورد نظر جمع‌آوری شد. هرچه طول دوره آماری بیشتر باشد، خطای برآورد کمتر و پیش‌بینی دقیق‌تر انجام می‌شود. باتوجه‌به کامل و در دسترس بودن اطلاعات همه پارامترهای هواشناسی مورد نظر در بازه زمانی مذکور، این دوره آماری انتخاب شد. جدول ۱، میانگین ۲۰ ساله داده‌های هواشناسی شامل باران، حداقل دما، حداکثر دما، میانگین دما، رطوبت نسبی، سرعت باد، تابش خورشیدی و اقلیم را در هریک از ایستگاه‌های مورد مطالعه نشان می‌دهد.

جدول ۱: میانگین پارامترهای هواشناسی ایستگاه‌های مورد مطالعه

ایستگاه	اقلیم (طبقه‌بندی کوپن)	حداقل دما (°C)	حداکثر دما (°C)	میانگین دما (°C)	رطوبت نسبی (%)	سرعت باد (m s <sup>-1</sup> )	میانگین بارندگی (mm year <sup>-1</sup> )	تابش خورشیدی (MJ m <sup>-2</sup> day <sup>-1</sup> )
چابهار	بیابانی گرم و خشک	۱۲/۵	۴۱/۸	۲۷/۱	۷۶	۲/۵	۱۱۴/۲۶	۱۰/۵
بندرعباس	بیابانی گرم میانه	۶/۴	۴۷/۵	۲۶/۹	۶۷	۲/۲	۱۶۴/۹۲	۹/۶
بوشهر	نیمه‌بیابانی گرم و خشک	۶/۹	۴۵/۳	۲۶/۱	۶۸	۲/۱	۲۰۱/۱۵	۸/۶
آبادان	بیابانی گرم شدید	۲/۳	۵۱/۶	۲۶/۹	۴۵/۴	۲/۸	۳۵۸/۴۱	۸/۵

## ۲.۲. برآورد تبخیر-تعرق مرجع

در این تحقیق برای محاسبه تبخیر-تعرق مرجع از روش فائو-پنمن-مانتیت استفاده شد. این روش در میان روش‌های مختلف تبخیر-تعرق گیاه مرجع بیشترین دقت را دارا بوده است [۲۳ و ۲۴]. از این رو مقادیر محاسبه‌شده توسط روش فائو-پنمن-مانتیت به‌عنوان مقدار استاندارد و هدف برای واسنجی و صحت‌سنجی سناریوهای مختلف روش‌های داده‌کاوی مورد استفاده قرار گرفت.

## معادله فائو-پنمن-مانتیت

$$ET_0 = \frac{0.408\Delta(Rn-G) + \gamma \frac{C_n}{T+273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d U_2)} \quad (1)$$

در این معادله،  $ET_0$  تبخیر-تعرق مرجع ( $mm \text{ day}^{-1}$ )،  $Rn$  تابش خالص ورودی به سطح گیاه ( $MJ \text{ m}^{-2} \text{ day}^{-1}$ )،  $G$  شار گرمای خاک ( $MJ \text{ m}^{-2} \text{ day}^{-1}$ )،  $T_{mean}$  میانگین دمای هوای روزانه ( $^{\circ}C$ )،  $U_2$  میانگین روزانه سرعت باد در ارتفاع دومتری از سطح زمین ( $m \text{ s}^{-1}$ )،  $e_s$  فشار بخار اشباع ( $KPa$ )،  $e_a$  فشار بخار واقعی ( $KPa$ )،  $es-ea$  کمبود فشار بخار اشباع ( $KPa$ )،  $\Delta$  شیب منحنی فشار بخار اشباع ( $KPa^{\circ}C^{-1}$ )،  $\gamma$  ضریب سایکرومتری ( $KPa^{\circ}C^{-1}$ )،  $C_n$  ضریبی برای گیاه مرجع که مقدار آن ۹۰۰ است ( $Kg^{\circ}K \text{ K}^{-1} \text{ day}^{-1}$ ) و  $C_d$  ضریب باد برای گیاه مرجع که مقدار آن ۰/۳۴ است ( $s \text{ m}^{-1}$ ).

## ۳.۲. روش‌های داده‌کاوی مورد مطالعه

بعد از محاسبه تبخیر-تعرق به روش مینا (فائو-پنمن-مانتیت)، از بین کل داده‌های موجود ۷۰ درصد برای آموزش و ۳۰ درصد برای آزمون روش‌های داده‌کاوی مورد استفاده قرار گرفتند. برای بررسی تأثیر متغیرهای مختلف هواشناسی (شامل پارامترهای دمای حداقل، دمای حداکثر، دمای میانگین، رطوبت نسبی، تابش خورشیدی و سرعت باد) و تعیین الگوهای ورودی به هر مدل از روش تجزیه به مؤلفه‌های اصلی ( $PCA^1$ ) از طریق همبستگی پارامترهای هواشناسی با تبخیر-تعرق مرجع استفاده شد. روش تحلیل مؤلفه‌های اصلی از خانواده بردارهای ویژه است که عمدتاً به‌عنوان یک روش ریاضی برای کاهش حجم داده‌ها به کار گرفته می‌شود. این روش، مجموعه متغیرهای اصلی را به مجموعه کوچک‌تری تبدیل می‌کند؛ به‌طوری‌که این مجموعه کوچک، بیشترین پراش موجود در داده‌ها را توجیه نماید. هدف از تحلیل مؤلفه‌های اصلی آن است که پراش موجود در داده‌های چندین متغیر را به مؤلفه‌هایی تبدیل می‌کند که اولین مؤلفه تا آنجا که ممکن است بیشترین پراش ممکن بعد از مؤلفه و تا آخر را توجیه اول نماید. به‌علاوه در این روش، هر مؤلفه مستقل از مؤلفه دیگر است و بین مؤلفه‌ها همبستگی وجود ندارد. اگر متغیرهای  $X_1, X_2, \dots, X_p$  را بررسی کنیم، تابع خطی  $PC_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$  را اولین مؤلفه اصلی گویند. به همین ترتیب دومین مؤلفه اصلی به‌صورت  $PC_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p$  خواهد بود و تا آخر. بنابراین به تعداد متغیرها، مؤلفه خواهیم داشت. اندازه هر کدام از مشاهدات در مؤلفه‌های جدید با نمرات مؤلفه‌ها محاسبه می‌شود. در نتیجه به‌جای اندازه مشاهدات، نمرات آن‌ها در مؤلفه‌های جدید مورد بررسی قرار می‌گیرد [۲۴].

سپس با استفاده از نتایج به‌دست‌آمده، از همبستگی ترکیب‌های مختلفی از داده‌های هواشناسی به‌عنوان ورودی مدل‌ها و مقادیر محاسبه‌شده متناظر به‌عنوان خروجی مدل‌ها استفاده شد و شبکه، آموزش داده شد. نتایج حاصل از این ترکیب‌ها مورد ارزیابی قرار گرفت و بهترین ترکیب برای برآورد تبخیر-تعرق در نظر گرفته شد. شایان ذکر است که برای اجرای مدل‌های داده‌کاوی مورد استفاده در تحقیق حاضر، از نرم‌افزار Weka استفاده شد.

## ۴.۲. ماشین بردار پشتیبان (SVR)

مفهوم اولیه ماشین‌های بردار پشتیبان، اولین بار توسط محقق روسی به نام واپنیک در سال ۱۹۹۵ مطرح شد. این مفهوم، روش جدیدی را بر مبنای طبقه‌بندی ارائه می‌دهد و یکی از ویژگی‌های آن، طراحی دسته‌بندی‌کننده با حداکثر تعمیم است [۲۵]. هدف ماشین‌های بردار پشتیبان، یافتن یک تابع پیوسته با مقدار حقیقی است که بتواند به‌ازای مقادیر ورودی، پیش‌بینی دقیقی از مقدار خروجی داشته باشد [۲۶]. به‌طور کلی ماشین بردار پشتیبان نوعی سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است [۲۷]. ماشین بردار پشتیبان، روشی است که هم برای گروه‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها در مسائل رگرسیون به کار می‌رود؛ به‌طوری‌که کمترین خطا در گروه‌بندی داده‌ها با تابع برازش رخ می‌دهد. SVM الگوریتمی است که نوع خاصی از مدل‌های خطی را می‌یابد که حداکثر حاشیه ابرصفحه را حاصل می‌کند. به نزدیک‌ترین نقاط آموزشی به حداکثر حاشیه ابرصفحه، بردارهای پشتیبان گفته می‌شود و فقط از این بردارها برای مشخص کردن مرز بین طبقات استفاده می‌شود. در یک مدل رگرسیونی SVM لازم است وابستگی تابع متغیر وابسته ( $y$ ) به مجموعه‌ای از متغیرهای مستقل ( $x$ ) تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای مستقل و وابسته توسط یک تابع معین  $f$  به‌علاوه یک مقدار اضافی نویز مشخص می‌شود [۲۸]:

$$y = f(x) + \text{noise} \quad (2)$$

بنابراین موضوع اصلی، پیدا کردن فرم اصلی تابع  $f$  است که بتواند به‌صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است، پیش‌بینی کند. در روش SVM  $\epsilon$  - از توابع هسته‌ای (کرنل) به‌منظور جداسازی صفحات و بهینه‌سازی فواصل موجود بین داده‌های مشاهده‌ای و تخمینی استفاده می‌شود؛ اما در مدل V-SVM از روش برنامه‌ریزی غیرخطی برای کاهش خطای پیش‌بینی استفاده می‌شود. در این مطالعه، مدل SVM  $\epsilon$  - به‌دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده شد. برای این مدل، تابع خطا به‌صورت زیر تعریف می‌شود [۲۸]:

$$C \sum_{i=1}^n \epsilon_i^* + C \sum_{i=1}^n \epsilon_i + \frac{1}{1} W^T W \quad (۳)$$

تابع خطای مذکور لازم است که باتوجه به محدودیت‌های زیر حداقل گردد:

$$W^T \phi(x_i) + b - y_i \leq \epsilon + \epsilon_i \quad (۴)$$

$$y_i - W^T \phi(x_i) + b \leq \epsilon + \epsilon_i \quad (۵)$$

$$\epsilon_i \times \epsilon^* \geq 0 \quad (۶)$$

در این روابط،  $C$  ثابت گنجایش،  $W$  بردار ضریب،  $W^T$  ترانزپوز بردار ضریب،  $\epsilon^*$  و  $\epsilon_i$  ضرایب کمبود،  $b$  ضریب ثابت،  $n$  الگوی آموزش مدل و  $\phi$  تابع کرنل است.

ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسئله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید باتوجه به این پارامترها تابع کرنل را انتخاب کرد که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل، کرنل خطی به کار برده می‌شود [۲۸].

## ۲.۵. درخت تصمیم (M5)

روش درخت تصمیم، زیرمجموعه‌ای از روش‌های یادگیری ماشینی و داده‌کاوی است. روش‌های یادگیری ماشینی و داده‌کاوی، قابلیت کشف نیمه‌اتوماتیک الگوهای ارزشمندی از داده‌ها را دارند. در سال‌های اخیر مدل درختی، تحول قابل توجهی در مسائل طبقه‌بندی و پیش‌بینی ایجاد نموده است. از جمله دلایل استفاده از روش درختی می‌توان به این موارد اشاره کرد که روش درختی به‌طور مستقیم با متغیرهای پیش‌بینی‌کننده مرتبط است؛ بنابراین نتایج روش برای فهم و شبیه‌سازی آسان هستند. همچنین درخت‌های تصمیم‌گیری غیرپارامتریک بوده و هیچ دخالتی از سوی کاربر روی آن‌ها صورت نمی‌گیرد. از طرف دیگر، خروجی روش از دقت بالایی برخوردار است که می‌توان آن را با سایر روش‌ها مقایسه کرد [۲۹]. درخت تصمیم برای انجام پیش‌بینی، ساختاری مشابه درخت ایجاد می‌کند؛ به این صورت که ابتدا کار خود را با استفاده از تمام نمونه‌های آموزشی شروع می‌کند و متغیری که بهترین دسته‌بندی را انجام می‌دهد، انتخاب می‌کند و زیرمجموعه‌هایی تشکیل می‌دهد. شاخه‌های درخت، نتیجه آزمونی است که در هر مرحله توسط الگوریتم روی گره‌های میانی صورت می‌گیرد. پیش‌بینی‌ها نیز در برگ‌های درخت ظاهر می‌شود [۲۹]. روش درختی، قابلیت پیش‌بینی متغیرهای پیوسته عددی از روی صفات عددی را دارد و نتایج پیش‌بینی‌شده به‌صورت مدل‌های رگرسیونی خطی چندمتغیره در برگ‌های درخت ظاهر می‌شود. معیار تقسیم در یک گره براساس انتخاب انحراف معیار مقادیر خروجی که به آن گره می‌رسند، به‌عنوان معیاری از خطاست. با آزمودن هر پارامتر در گره کاهش مورد انتظار در خطا محاسبه می‌شود. خطای مدل عموماً با اندازه‌گیری دقت پیش‌بینی مقادیر هدف موارد دیده‌نشده سنجش می‌شود [۳۰]. فرمول محاسبه کاهش انحرافات (SDR) به‌صورت روابط (۷) و (۸) هستند:

$$SDR = Sd(T) - \sum_{i=1}^N \frac{|Ti|}{|T|} Sd(Ti) \quad (۷)$$

$$Sd(T) = \sqrt{\frac{1}{N} (\sum_{i=1}^N Yi^2 - \frac{1}{N} (\sum_{i=1}^N Yi)^2)} \quad (۸)$$

در این روابط، SDR کاهش انحراف معیار و  $T$  مجموعه‌ای از نمونه‌هاست که به هر گره وارد می‌شود.  $Ti$  نشان‌دهنده زیرمجموعه‌ای از نمونه‌هاست که آمین نتیجه آزمون پتانسیلی را دارند.  $Sd$  انحراف معیار،  $Yi$  مقدار عددی ویژگی هدف نمونه  $i$  و  $N$  شماره داده‌ها را نشان می‌دهد [۲۹].

## ۲.۶. جنگل تصادفی

جنگل تصادفی، یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی مکرر از هر پدیده ترکیب می‌کند. جنگل تصادفی می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیرخطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد. همچنین

می‌تواند انواع مختلف داده‌ها را به‌علت نبود توزیع نرمال داده‌های استفاده‌شده، در تجزیه و تحلیل بگنجانند و ترکیب کند. مدل جنگل تصادفی برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با سرعت بالایی انجام می‌دهد و برخلاف مدل‌های کلاسیک چون رگرسیون که فقط بر پایه یک مدل تکیه دارند، با استفاده از صدها و هزاران درخت از اطلاعات بیشتری در داده‌ها استفاده می‌کند تا بتوان استنباط بهتری از متغیرها داشت. این روش، یک روش مدل ناپارامتری و متعلق به خانواده روش‌های دسته‌جمعی است [۳۱]. جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن  $n$  داده مشاهده‌ای مستقل ساخته می‌شود:

$$(Y_i, X_i), \quad i = 1, \dots, n \quad (9)$$

این الگوریتم از جمله دسته‌بندی‌هایی است که روش Bagging را استفاده می‌کند. این روش، ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت‌استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به‌طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت‌استرپ به تعداد زیاد (برای مثال ۲۰۰۰ مرتبه) نمونه‌های  $n$  تایی از مجموعه داده‌های مشاهده‌ای اولیه، نمونه‌برداری می‌شوند. در طی فرایند نمونه‌گیری، حدود یک‌سوم از داده‌ها نمونه‌گیری نمی‌شوند و برای نمونه خارج از کیسه (از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد ناریب خطا استفاده می‌شود) در نظر گرفته می‌شوند. سپس روی هر نمونه بوت‌استرپ یک درخت گسترش داده می‌شود. طی فرایند ساخت درخت در هر شاخه، از بین تمام  $M$  متغیر مستقل به‌صورت تصادفی  $m$  متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی، نسبت  $m/M$  برابر یک‌سوم است و برای کلاسه‌بندی برابر با  $m = \sqrt{M}$  پیشنهاد شده است. پس از ساخت تمام درخت داده‌های آزمون به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها، مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش جنگل تصادفی زمانی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد، یک روش پیش‌بینی کارآمد است [۳۲].

## ۲.۷. رگرسیون فرایند گاوسی (GPR)

رگرسیون فرایند گاوسی، مجموعه‌ای از متغیرهای تصادفی است که تعداد محدودی از آن‌ها با توزیع‌های گاوسی سازگار شده‌اند. این فرایند به‌طور کامل توسط تابع میانگین و تابع کواریانس معین می‌شود. این فرایند تعمیمی طبیعی از توزیع گاوسی است که میانگین و کواریانس آن به‌ترتیب یک بردار و یک ماتریس است. مدل‌های رگرسیون فرایند گاوسی بر مبنای این فرض هستند که داده‌ها باید حاوی اطلاعاتی درباره همدیگر باشند. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی بوده، درحالی‌که فرایند گاوسی بیانگر توزیع بین توابع است؛ درنتیجه مدل‌های فرایند گاوسی به‌دلیل دانش قبلی درباره وابستگی‌های تابعی و داده‌ها برای تعمیم به هیچ فرایند اعتبارسنجی نیاز ندارند [۳۳]. فرایند گاوسی  $f(x)$  توسط توابع میانگین  $m(x)$  و کواریانس به‌شکل زیر تعریف می‌شود [۳۴]:

$$m(x) = E(f(x)) \quad (10)$$

$$k(x, x') = E((f(x) - m(x))(f(x') - m(x'))) \quad (11)$$

که در آن،  $k(x, x')$  تابع کرنل بوده که در نقاط  $x$  و  $x'$  محاسبه می‌شود. فرایند گاوسی  $f(x)$  به‌صورت زیر بیان می‌شود:

$$f(x) = GP(m(x), k(x, x')) \quad (12)$$

برای ساده‌سازی، مقدار تابع میانگین صفر می‌شود [۳۵].

## ۲.۸. آماره‌های ارزیابی

در این تحقیق برای ارزیابی صحت و اعتبار مدل‌های مورد استفاده از شاخص‌های آماری ریشه میانگین مربعات خطا (RMSE)، ضریب تبیین ( $R^2$ )، و میانگین مطلق خطا (MAE) استفاده شد (رابطه‌های ۱۳ تا ۱۵).

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - x_i^*)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_i^*)^2} \quad (14)$$

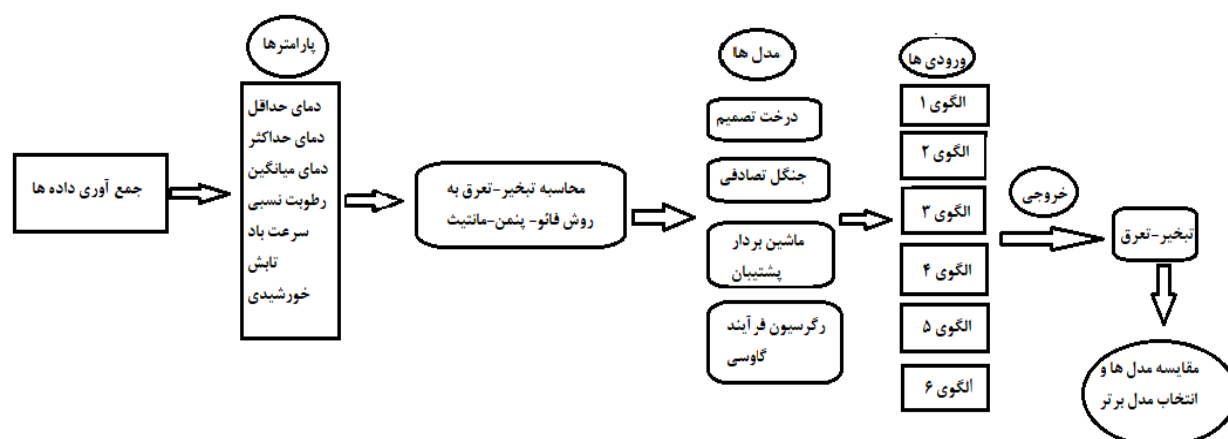
$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x_i^*| \quad (15)$$

در این روابط  $x_i$  مقدار اندازه‌گیری‌شده،  $x_i^*$  مقادیر برآوردشده و  $n$  تعداد نمونه است.

شایان ذکر است که هرچه ضریب  $R^2$  به یک نزدیک‌تر و معیارهای RMSE و MAE به صفر نزدیک‌تر باشند، مدل از اعتبار بیشتری برخوردار است.



شکل ۲ نیز مراحل انجام تحقیق را نشان می‌دهد.



شکل ۲: الگوریتم مورد استفاده در تحقیق

### ۳. نتایج و بحث

همان‌طور که در بخش‌های قبل توضیح داده شد، مدل‌های داده‌محور، نیازی به معادلات با پایه‌های فیزیکی، که بیشتر به صورت غیرخطی و پیچیده‌اند، ندارند؛ ولی درک فرایندها و تأثیر عوامل مورد نظر در داده‌های ورودی روی تولید خروجی آن‌ها بسیار مهم است و در صورت درک نکردن آن، نمی‌توان انتظار نتایج مطلوبی داشت. لذا برای مدل‌سازی، پارامترهای هواشناسی مؤثر در تبخیر-تعرق مرجع، میزان تأثیر هر یک از پارامترها با استفاده از روش تجزیه به مؤلفه‌های اصلی از طریق مقادیر همبستگی پارامترها با میزان تبخیر-تعرق مرجع محاسبه شد که نتایج در جدول ۲ ارائه شده است. نتایج نشان داد پارامترهای دمای حداقل، دمای حداکثر و دمای میانگین در تمام ایستگاه‌های مورد مطالعه، بیشترین همبستگی و سرعت باد، همبستگی کمتری با مقادیر تبخیر-تعرق مرجع داشتند. رطوبت نسبی همبستگی منفی با مقادیر تبخیر-تعرق مرجع داشت. شایان ذکر است که مقادیر ضرایب مثبت همبستگی نشان‌دهنده اثر همسوی تغییرات هر کدام از متغیرها با تبخیر-تعرق مرجع و مقادیر منفی نیز نشان‌دهنده تأثیر معکوس تغییرات پارامترها بر میزان تبخیر-تعرق مرجع است. نتایج به‌دست‌آمده از مطالعات سایر محققان نیز نشان داده است که در بین عوامل هواشناسی مؤثر در تبخیر-تعرق، متغیرهای دما، ساعات آفتابی و تابش خورشیدی بیشترین تأثیر را دارند [۳۶]. در تحقیقی بیان شد دمای حداقل روزانه و رطوبت نسبی به صورت معکوس و دمای حداکثر و متوسط دمای روزانه، به صورت مستقیم بر روی تبخیر تأثیرگذار است و بین متغیرهای میانگین دما، رطوبت نسبی هوا و مؤلفه تبخیر، همبستگی بالایی مشاهده شد [۳۷].

جدول ۲: تجزیه به مؤلفه‌های اصلی براساس مقادیر همبستگی پارامترها با تبخیر-تعرق مرجع در ایستگاه‌های مورد مطالعه

پارامترهای هواشناسی	چابهار	بندرعباس	بوشهر	آبادان
حداقل دما (°C)	۰/۸۹	۰/۹۱	۰/۹۲	۰/۸۶
حداکثر دما (°C)	۰/۹۲	۰/۹۳	۰/۹۴	۰/۸۹
میانگین دما (°C)	۰/۸۳	۰/۸۷	۰/۸۶	۰/۸۴
رطوبت نسبی (/)	-۰/۶۸	-۰/۶۲	-۰/۶۵	-۰/۶۱
سرعت باد (m s <sup>-1</sup> )	۰/۴۵	۰/۳۹	۰/۳۸	۰/۴۴
تابش خورشیدی (MJ m <sup>-2</sup> day <sup>-1</sup> )	۰/۵۷	۰/۵۲	۰/۵۴	۰/۵۱

برای مدل‌سازی مقادیر تبخیر-تعرق مرجع و بررسی کارایی روش‌های داده‌کاوی ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و رگرسیون فرآیند گاوسی از نتایج جدول ۲ استفاده شد. برای این منظور، معیارهای خطای ارزیابی نتایج برای چهار روش مذکور در ۶ سناریو محاسبه شد. بر این اساس در سناریوی ششم، تمام متغیرها در مدل‌سازی مورد استفاده قرار گرفتند. در سناریوهای بعدی متغیری که میزان همبستگی کمتری را با مقادیر تبخیر-تعرق مرجع داشت، در مدل‌سازی حذف شد. این فرآیند به همین ترتیب ادامه یافت و در نهایت در سناریوی اول، فقط از یک متغیر که بیشترین همبستگی را با مقادیر تبخیر-تعرق مرجع دارد، استفاده شد. جدول ۳ الگوهای مختلف ورودی به مدل‌ها را نشان می‌دهد.

جدول ۳: الگوهای مختلف ورودی، برای تخمین مقدار تبخیر-تعرق مرجع در مدل‌های مورد مطالعه

سنا ریبو	پارامترهای ورودی	خروجی
۱	حداکثر دما	تبخیر-تعرق مرجع
۲	حداقل دما	تبخیر-تعرق مرجع
۳	میانگین دما	تبخیر-تعرق مرجع
۴	میانگین دما	تبخیر-تعرق مرجع
۵	میانگین دما	تبخیر-تعرق مرجع
۶	میانگین دما	تبخیر-تعرق مرجع

پس از طراحی سناریوها برای مدل‌های مختلف، خروجی این مدل‌ها دریافت و با شاخص‌های آماری مورد ارزیابی قرار گرفت. جدول ۴ نتایج ارزیابی را برای چهار روش داده‌کاوی (شامل ماشین بردار پشتیبان، جنگل تصادفی، درخت تصمیم و رگرسیون فرایند گاوسی) نشان می‌دهد.

جدول ۴: مقادیر شاخص‌های آماری ارزیابی روش‌های داده‌کاوی مورد مطالعه

ایستگاه	مدل	ماشین بردار پشتیبان			جنگل تصادفی			درخت تصمیم			رگرسیون فرایند گاوسی		
		R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
چابهار	۱	۰/۸۸	۱/۹۶	۱/۶۱	۰/۹۲	۱/۳۷	۰/۸۴	۰/۸۹	۱/۵۹	۱/۲۴	۰/۷۸	۱/۱۸	۰/۹۱
	۲	۰/۸۹	۱/۸۵	۱/۴۲	۰/۹۴	۱/۲۶	۰/۷۹	۰/۹۲	۱/۴۷	۱/۱۱	۰/۷۳	۱/۱۱	۰/۹۴
	۳	۰/۹۱	۱/۴۹	۱/۲۲	۰/۹۴	۱/۱۸	۰/۷۱	۰/۹۲	۱/۳۸	۰/۹۲	۰/۶۲	۰/۹۷	۰/۹۶
	۴	۰/۹۱	۱/۴۵	۰/۹۷	۰/۹۵	۱/۱۴	۰/۶۷	۰/۹۳	۱/۲۴	۰/۹۱	۰/۶۱	۰/۹۴	۰/۹۶
	۵	۰/۹۱	۱/۸۹	۱/۵۷	۰/۹۳	۱/۲۹	۰/۸۳	۰/۹۱	۱/۵۶	۱/۲۳	۰/۸۱	۱/۱۵	۰/۹۲
	۶	۰/۹۲	۱/۳۶	۰/۸۶	۰/۹۳	۰/۹۸	۰/۵۹	۰/۹۲	۱/۱۵	۰/۷۸	۰/۵۶	۰/۸۶	۰/۹۴
بندرعباس	۱	۰/۸۶	۲/۴۵	۱/۸۳	۰/۹۱	۲/۰۲	۱/۸۶	۰/۹۱	۱/۷۹	۱/۸۸	۱/۹۳	۱/۸۵	۰/۹۲
	۲	۰/۸۸	۲/۳۷	۱/۷۹	۰/۹۱	۱/۹۶	۱/۷۵	۰/۹۱	۱/۶۷	۱/۷۵	۱/۳۱	۱/۵۵	۰/۹۴
	۳	۰/۸۹	۲/۱۹	۱/۷۵	۰/۹۲	۱/۷۵	۱/۵۸	۰/۹۳	۱/۶۶	۱/۷۳	۱/۳۸	۱/۵۶	۰/۹۳
	۴	۰/۹۱	۱/۸۹	۱/۶۸	۰/۹۵	۱/۳۸	۱/۲۶	۰/۹۳	۱/۶۲	۱/۴۴	۱/۱۴	۱/۰۲	۰/۹۵
	۵	۰/۹۱	۱/۹۶	۱/۶۹	۰/۹۳	۱/۵۴	۱/۳۹	۰/۹۲	۱/۷۲	۱/۵۲	۱/۲۷	۱/۴۲	۰/۹۴
	۶	۰/۸۹	۲/۳۶	۱/۷۴	۰/۹۲	۱/۸۱	۱/۶۱	۰/۹۱	۱/۸۲	۱/۷۷	۱/۴۲	۱/۶۲	۰/۹۳
بوشهر	۱	۰/۸۳	۰/۸۹	۱/۹۲	۰/۹۲	۱/۳۷	۱/۲۸	۰/۸۴	۱/۸۴	۱/۶۷	۱/۳۴	۱/۲۵	۰/۹۱
	۲	۰/۸۷	۱/۷۶	۱/۶۵	۰/۹۲	۱/۴۵	۱/۱۴	۰/۸۹	۱/۲۷	۱/۲۱	۰/۸۶	۱/۱۳	۰/۹۳
	۳	۰/۸۹	۱/۳۲	۱/۴۲	۰/۹۳	۰/۹۷	۰/۸۵	۰/۹۳	۱/۱۵	۰/۹۹	۰/۷۵	۰/۹۵	۰/۹۴
	۴	۰/۸۷	۱/۸۲	۱/۹۵	۰/۹۵	۰/۸۱	۰/۷۶	۰/۹۴	۰/۹۵	۰/۸۵	۰/۶۹	۰/۷۴	۰/۹۷
	۵	۰/۹۱	۱/۱۴	۱/۲۳	۰/۹۴	۱/۳۹	۰/۹۶	۰/۹۱	۱/۰۶	۱/۵۶	۱/۱۳	۰/۸۴	۰/۹۵
	۶	۰/۸۵	۲/۰۳	۱/۸۱	۰/۹۱	۱/۶۲	۱/۰۵	۰/۹۱	۱/۲۳	۱/۹۴	۱/۳۲	۱/۱۱	۰/۹۴
آبادان	۱	۰/۸۵	۲/۱۸	۱/۴۴	۰/۹۴	۱/۶۱	۱/۲۶	۰/۸۹	۱/۷۸	۱/۳۳	۰/۸۸	۱/۴۶	۰/۹۲
	۲	۰/۸۷	۱/۸۶	۱/۱۹	۰/۹۴	۱/۴۷	۰/۹۷	۰/۹۱	۱/۵۴	۱/۰۸	۰/۶۶	۱/۰۵	۰/۹۴
	۳	۰/۸۴	۲/۳۲	۱/۶۳	۰/۹۴	۱/۱۷	۰/۷۴	۰/۹۳	۱/۲۵	۰/۸۳	۰/۵۹	۰/۸۷	۰/۹۵
	۴	۰/۸۹	۱/۶۴	۰/۹۴	۰/۸۹	۱/۷۵	۱/۳۵	۰/۸۷	۱/۹۵	۱/۴۲	۰/۷۲	۰/۹۷	۰/۹۲
	۵	۰/۸۴	۲/۴۱	۱/۷۱	۰/۹۱	۱/۸۳	۱/۴۲	۰/۸۹	۲/۱۵	۱/۵۳	۰/۸۱	۱/۲۶	۰/۹۱
	۶	۰/۸۲	۲/۷۴	۱/۹۶	۰/۸۸	۲/۱۲	۱/۷۶	۰/۸۶	۲/۴۱	۱/۸۲	۰/۸۶	۱/۶۹	۰/۹۱

نتایج ارزیابی روش‌های داده‌کاوی (جدول ۴) نشان داد هر چهار روش داده‌کاوی در مناطق مورد مطالعه به‌خوبی توانسته‌اند مقادیر تبخیر-تعرق مرجع را برآورد کنند. در هر چهار ایستگاه، روش رگرسیون فرایند گاوسی با داشتن بالاترین مقدار R<sup>2</sup> و کمترین مقادیر RMSE و MAE برآورد بهتری از مقادیر تبخیر-تعرق مرجع داشتند و روش‌های جنگل تصادفی، درخت تصمیم و ماشین بردار پشتیبان به ترتیب در رتبه‌های بعدی قرار گرفتند. محققان بیان داشتند از بین روش‌های داده‌کاوی ماشین بردار پشتیبان، رگرسیون فرایند گاوسی و نزدیک‌ترین

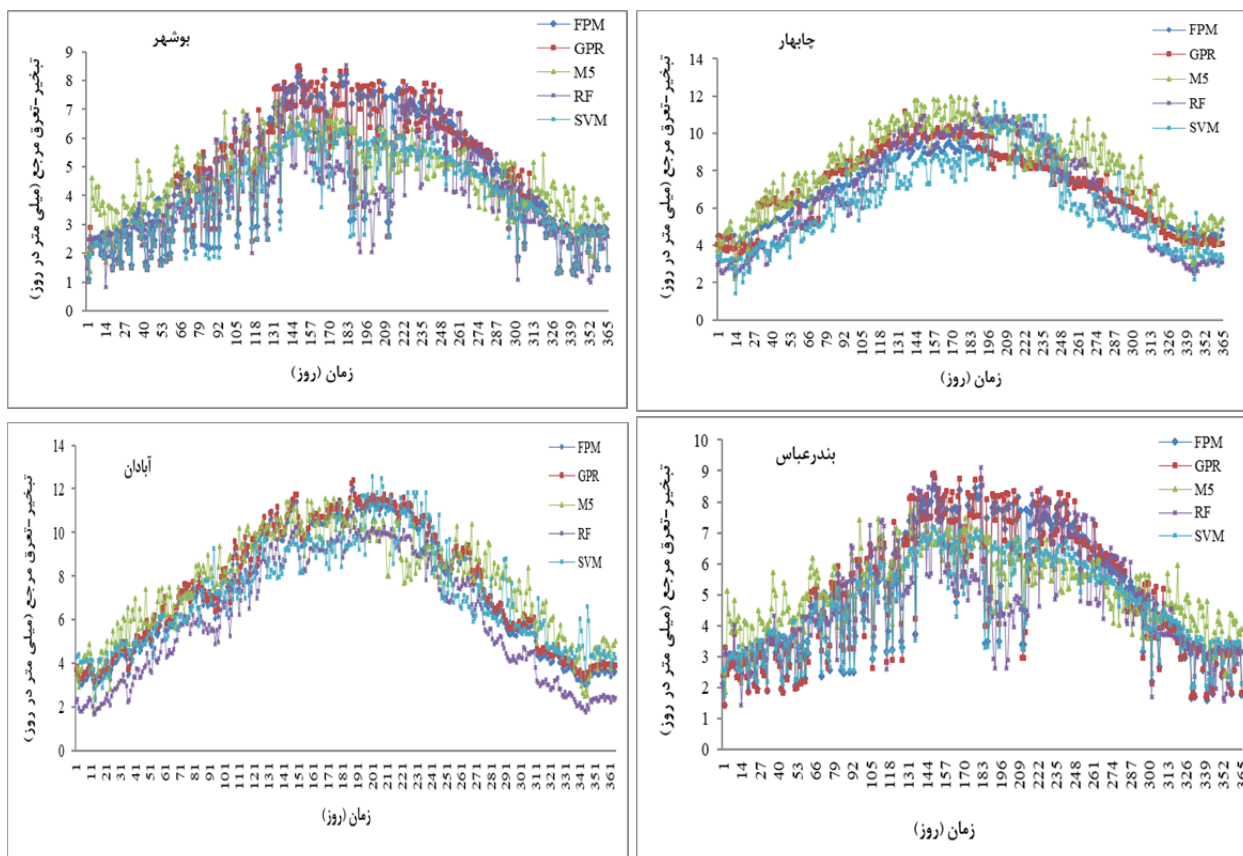
همسایگی (IBK)، در ایستگاه‌های تبریز، سراب و هریس، روش ماشین بردار پشتیبان و در ایستگاه‌های بستان‌آباد و مراغه، روش رگرسیون فرایند گاوسی در مقایسه با سایر روش‌ها دقت بیشتری داشته‌اند [۳۸]. در مطالعه دیگری نیز بیان شد رگرسیون فرایند گاوسی، بهترین برآورد در تبخیر از تشت را در منطقه گلستان داشت [۳۹].

در ایستگاه چابهار، روش رگرسیون فرایند گاوسی در الگوی ۶ با بیشترین  $R^2$  (۰/۹۴) و کمترین مقادیر RMSE و MAE به‌ترتیب برابر ۰/۸۶ و ۰/۵۶ برآورد بهتری از تبخیر-تعرق مرجع داشتند. سایر روش‌ها نیز در الگوی ۶ برآورد بهتری از تبخیر-تعرق مرجع دارا بودند. در ایستگاه بندرعباس همه روش‌ها در الگوی ۴، بهترین برآورد تبخیر-تعرق مرجع را داشتند. در ایستگاه بوشهر، روش‌های رگرسیون فرایند گاوسی، درخت تصمیم و جنگل تصادفی در الگوی ۴ و روش ماشین بردار پشتیبان در الگوی ۵، برآورد بهتری از تبخیر-تعرق مرجع داشتند. در ایستگاه آبادان، روش‌های رگرسیون فرایند گاوسی، درخت تصمیم و جنگل تصادفی در الگوی ۳ و روش ماشین بردار پشتیبان در الگوی ۴، برآورد بهتری از تبخیر-تعرق مرجع داشتند.

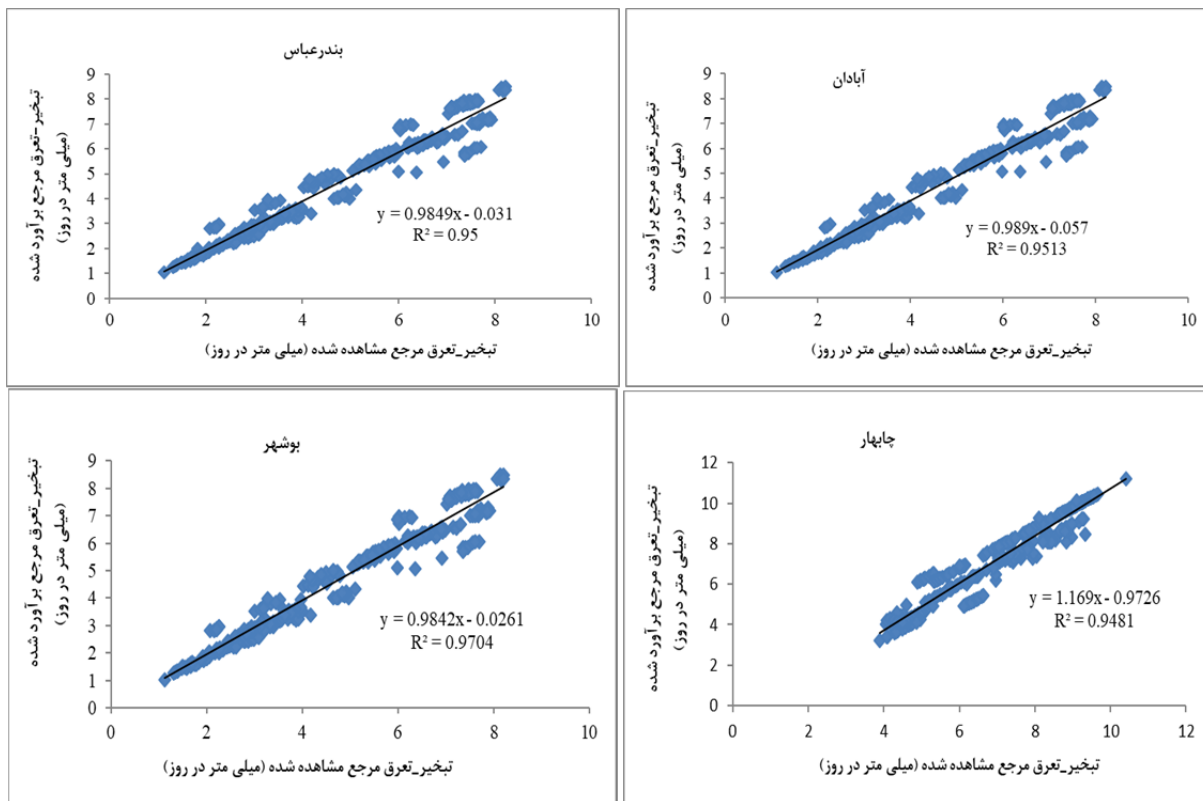
همچنین مقایسه بین مدل‌ها (جدول ۴) نشان داد در ایستگاه چابهار، ترکیب هر شش پارامتر هواشناسی (پارامترهای دمای حداقل، دمای حداکثر، دمای میانگین، رطوبت نسبی، تابش خورشیدی و سرعت باد)، برآورد بهتری از تبخیر-تعرق مرجع داشتند. حذف هر یک از پارامترهای هواشناسی باعث افزایش خطاهای RMSE و MAE و کاهش ضریب تبیین  $R^2$  شد. در ایستگاه بندرعباس، در هر چهار روش داده‌کاوی حذف سرعت باد و تابش خورشیدی باعث افزایش ضریب تبیین  $R^2$  و کاهش آماره‌های خطای RMSE و MAE شد. در ایستگاه بوشهر نیز در سه روش درخت تصمیم، جنگل تصادفی و رگرسیون فرایند گاوسی با حذف سرعت باد و تابش خورشیدی، ضریب تبیین  $R^2$  افزایش و آماره‌های خطای RMSE و MAE کاهش یافت. اما روش ماشین بردار پشتیبان در این ایستگاه فقط با حذف سرعت باد عملکرد بهتری در برآورد داشت. در ایستگاه آبادان، ترکیب پارامترهای دمای حداقل، دمای حداکثر و دمای میانگین (الگوی ۳)، برآورد بهتری از تبخیر-تعرق را در سه روش درخت تصمیم، جنگل تصادفی و رگرسیون فرایند گاوسی نشان داد؛ اما در این ایستگاه در روش ماشین بردار پشتیبان، اضافه نمودن رطوبت نسبی به ترکیب مذکور (الگوی ۴) تبخیر-تعرق را بهتر برآورد نمود. تخمین تبخیر-تعرق مرجع روزانه در بوشهر نشان داد که بیشترین حساسیت روش‌های هوش مصنوعی به پارامترهای دمای هواست [۱۵]. در تحقیقی دیگر مشاهده شد با استفاده از روش‌های هوش مصنوعی و دو پارامتر دمای حداکثر و دمای حداقل، تبخیر-تعرق مرجع به‌خوبی برآورد شد [۴۰]. با توجه به نتایج به‌دست‌آمده از تحقیق حاضر می‌توان یک الگوی مشخصی از داده‌های هواشناسی برای برآورد تبخیر-تعرق مرجع با استفاده از روش‌های داده‌کاوی در ایستگاه‌های نوار ساحلی جنوب ایران معرفی کرد.

در شکل (۳)، روند تغییرات روزانه تبخیر-تعرق مرجع محاسبه‌شده با روش فائو-پنمن-مانتیت و پیش‌بینی‌شده با روش‌های داده‌کاوی ارائه شده است. همان‌طور که در شکل مشاهده می‌شود، انطباق زیادی بین داده‌های محاسبه‌شده با روش فائو-پنمن-مانتیت و روش‌های داده‌کاوی وجود دارد. بنابراین می‌توان از این روش‌ها برای برآورد تبخیر-تعرق مرجع در زمانی که داده‌های محدودی در اختیار است، استفاده کرد.

به‌منظور ارزیابی بهتر کارایی روش رگرسیون فرایند گاوسی، نحوه پراکنش مقادیر پیش‌بینی‌شده تبخیر-تعرق مرجع تحت بهترین سناریو حول خط نیم‌ساز و مقایسه آن با مقادیر تبخیر-تعرق مرجع محاسبه‌شده با روش فائو-پنمن-مانتیت، در شکل (۴) ارائه شده است. میزان پراکندگی نقاط پیش‌بینی حاصل از روش رگرسیون فرایند گاوسی حول خط نیم‌ساز نشان‌دهنده دقت پیش‌بینی روش است. به‌طوری‌که هرچقدر میزان پراکنش از خط نیم‌ساز بیشتر باشد، اختلاف مقادیر پیش‌بینی‌شده با مقادیر تبخیر-تعرق مرجع محاسبه‌شده با روش فائو-پنمن-مانتیت بیشتر بوده و دقت مدل کمتر است. همان‌طور که مشاهده می‌شود، مدل رگرسیون فرایند گاوسی، در هر چهار ایستگاه، کمترین پراکنش را حول نیم‌ساز داشته و دارای بیشترین دقت و کمترین خطا بوده است و مدل به‌خوبی توانسته مقادیر تبخیر-تعرق مرجع را برآورد نماید. با توجه به معادله رگرسیونی به‌دست‌آمده از برازش داده‌های تبخیر-تعرق برآوردشده هر ایستگاه با داده‌های محاسبه‌شده فائو-پنمن-مانتیت (شکل ۳) می‌توان گفت مدل رگرسیون فرایند گاوسی مقادیر تبخیر-تعرق مرجع را کمتر از مقادیر محاسبه‌شده فائو-پنمن-مانتیت برآورد کرده است. دیگر محققان نیز در تحقیق خود به نتایج مشابهی دست یافتند و بیان داشتند مدل رگرسیون فرایند گاوسی به‌خوبی توانسته است مقادیر تبخیر از تشت را در ایستگاه تبریز برآورد نماید. ایشان همچنین بیان کردند روش‌های داده‌کاوی بررسی‌شده حساسیت زیادی به پارامترهای دمایی و تابش خورشیدی داشتند [۱۶].



شکل ۳: مقایسه روند تغییرات روزانه تبخیر-تعرق مرجع محاسبه شده با روش فائو-پنمن-مانتیت و روش های داده کاوی مورد مطالعه



شکل ۴: پراکنش مقادیر پیش بینی شده تبخیر-تعرق مرجع تحت بهترین سناریو روش رگرسیون فرایند گاوسی حول خط نیم سازه و مقایسه با مقادیر تبخیر-تعرق محاسبه شده از روش فائو-پنمن-مانتیت

#### ۴. نتیجه‌گیری

آگاهی از نیاز آبی گیاهان یکی از ملاحظات مهم در اصلاح بهره‌وری آب در کشاورزی است؛ از این رو پیش‌بینی تبخیر-تعرق مرجع برای محققان از اهمیت خاصی برخوردار است. برآورد دقیق تبخیر-تعرق مرجع از مهم‌ترین عوامل ارتقای مدیریت آب است. این موضوع در مناطق با نرخ تبخیر-تعرق بالا مانند نوار ساحلی جنوب کشور بیش از پیش اهمیت دارد. روش فائو-پنمن-مانتیت به‌عنوان فرمول استاندارد برآورد تبخیر-تعرق مرجع مورد استفاده قرار می‌گیرد؛ ولی این روش دارای محاسبات طولانی و پیچیده بوده و نیاز به اطلاعات زیادی دارد. از سوی دیگر به‌دلیل اینکه در برآورد تبخیر-تعرق روزانه تغییرات کوتاه‌مدت پارامترهای هواشناسی مؤثر بر تبخیر-تعرق در نظر گرفته می‌شود، انتخاب پارامترهای مؤثر و در صورت امکان حذف برخی از آن‌ها برای کاهش متغیرهای ورودی، می‌تواند کمک زیادی در برآورد سریع و کم‌هزینه تبخیر-تعرق روزانه به‌خصوص مدل‌سازی کند. استفاده از روش‌های داده‌کاوی با استفاده از داده‌های هواشناسی کمتر می‌تواند تبخیر-تعرق مرجع روزانه را برآورد نماید. در این تحقیق از روش‌های داده‌کاوی ماشین بردار پشتیبان، جنگل تصادفی، درخت تصمیم و رگرسیون فرایند گاوسی با شش الگوی مختلف برای برآورد تبخیر-تعرق مرجع روزانه چهار ایستگاه نوار ساحلی جنوب کشور (چابهار، بندرعباس، بوشهر و آبادان) استفاده شد. همان‌طور که نتایج نشان دادند، همبستگی بالایی بین تبخیر-تعرق برآوردشده از روش‌های داده‌کاوی با روش فائو-پنمن-مانتیت وجود دارد و می‌توان از این روش‌ها برای برآورد تبخیر-تعرق مرجع روزانه در زمانی که داده‌های محدودتری در دسترس است، استفاده کرد. از بین روش‌های مختلف، روش رگرسیون فرایند گاوسی با دارا بودن کمترین آماره‌های خطا و بیشترین ضریب تبیین، برآورد بهتری از تبخیر-تعرق مرجع داشت و روش‌های جنگل تصادفی، درخت تصمیم و ماشین بردار پشتیبان به‌ترتیب در رتبه‌های بعدی قرار گرفتند. از بین الگوهای مورد بررسی در چابهار الگوی ۶، در بندرعباس و بوشهر الگوی ۴، در آبادان الگوی ۳ بهترین برآورد را داشتند. علت تفاوت در نتیجه به‌دست‌آمده در هر ایستگاه، شاید به‌دلیل وضعیت آب‌وهوایی متفاوت هر ایستگاه باشد. در ایستگاه چابهار، اقلیم منطقه بیابانی گرم و خشک، ایستگاه بندرعباس، اقلیم منطقه بیابانی گرم میانه، ایستگاه بوشهر، اقلیم منطقه نیمه‌بیابانی گرم و خشک و ایستگاه آبادان، اقلیم منطقه بیابانی گرم شدید است. به نظر می‌رسد با توجه به اینکه ایستگاه آبادان نسبت به سه ایستگاه دیگر از دریا فاصله دارد و رطوبت نسبی کمتری نسبت به سه ایستگاه دیگر دارد، تأثیر این پارامتر بر مقدار تبخیر-تعرق کمتر است. همچنین شدت تابش خورشیدی نیز در این ایستگاه کمتر بود. به همین دلیل، الگوی ۳ شامل ترکیب پارامترهای دمای حداقل، دمای حداکثر و دمای میانگین، برآورد بهتری از تبخیر-تعرق داشت. در ایستگاه‌های بندرعباس و بوشهر نیز به‌دلیل نزدیکی اقلیم دو منطقه و وضعیت آب‌وهوایی، نتیجه یکسانی حاصل شد. در ایستگاه چابهار به‌دلیل بالا بودن پارامترهای هواشناسی مورد مطالعه نسبت به سایر ایستگاه‌ها، نتیجه متفاوت‌تری به دست آمد و الگوی ۶ با ترکیب همه پارامترهای هواشناسی برآورد بهتری داشت. از بین پارامترهای هواشناسی، پارامترهای حداکثر دما، حداقل دما و میانگین دما بیشترین تأثیر را بر تبخیر-تعرق داشت. با توجه به نقش مهم تبخیر-تعرق مرجع در مطالعات کشاورزی، طرح‌های مدیریت منابع آب، تعیین نیاز آبی و تعیین هیدرومدول برای طراحی شبکه‌های آبیاری و زهکشی، نتایج حاصل از این تحقیق می‌تواند راهکار مناسبی برای محاسبه و برآورد این پارامتر مهم هیدرولوژیکی به مدیران و طراحان ارائه دهد؛ به‌ویژه که هر چهار مدل مورد استفاده با اطلاعات ورودی کمتر نسبت به روش فائو-پنمن-مانتیت عملکرد خوبی داشته‌اند. در پایان پیشنهاد می‌شود عملکرد مدل‌های مورد استفاده در این تحقیق نسبت به داده‌های لایسیمیتری ارزیابی شود. همچنین با توجه به سادگی و کاربردی بودن استفاده از روش‌های داده‌کاوی مورد مطالعه، به‌منظور برآورد تبخیر-تعرق مرجع، در اقلیم‌هایی مشابه با اقلیم مناطق مورد مطالعه، این روش‌ها پیشنهاد می‌شوند. لازم است روش‌های مورد استفاده در مطالعه حاضر در اقلیم‌های مختلف مورد بررسی قرار گیرد و برای هر اقلیم، بهترین روش مشخص گردد.

#### سپاسگزاری

این پژوهش با حمایت مالی دانشگاه زابل کد پژوهانه IR-UOZ-GR-1837 انجام شده است.

## منابع

- [1] Maeda EE, Wiberg DA, Pellikka PKE. Estimating reference evapotranspiration using sensing empirical models in a region with limited data availability in Kenya. *Applied Geography*. 2010; 31(2011): 251-258.
- [2] Bos MG, Kselik RG, Allen k, Molden DJ. *Water Requirements for Irrigation and the Environment*. Springer, 2009. p. 186.
- [3] Panahi S, Rezvanizadeh F, Samadianfard S. Evaluation and comparison of experimental methods for estimation of reference evapotranspiration in Tabriz station. *The first international conference on Iran's natural hazards and environmental crises, solutions and challenges*, 2016; 9 p. [Persian]
- [4] Alizadeh A. *Irrigation System Design*. Ferdowsi University of Mashhad, 2002. 450pp. [Persian].
- [5] Chiew FHS, Kamaladassa NN, Malano HM, MacMahon, TA. 1995. Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agricultural Water management*, 28: 9- 21.
- [6] Hashemi E, Ahmadpari H, Kohneh K. Comparison of different methods of estimating potential evapotranspiration with FAO Penman-Monteith method (Case study: Sepidan region). *Nivar*. 2017; 41, 13-22. DOI: 10.30467/nivar.2017.51886. [In Persian]
- [7] Samadianfard S, Panahi S. Estimating Daily Reference Evapotranspiration using Data Mining Methods of Support Vector Regression and M5 Model Tree. *Journal of Watershed Management Research*. 2018; 18(10): 157-167. [Persian].
- [8] Wang Z, Xie P, Lai C, Chen X, Wu X, Zeng Z, Li J. Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961– 2013. *Journal of Hydrology*. 2017; 544: 97-108.
- [9] Ansari Ghojghar M, Pourgholam-Amiji M, Bazrafshan J, Liaghat A, Araghinejad S. Performance Comparison of Statistical, Fuzzy and Perceptron Neural Network Models in Forecasting Dust Storms in Critical Regions in Iran. *Iranian Journal of Soil and Water Research*. 2020; 51(8): 2051-2063. [Persian].
- [10] Pai P, Hong WCA. Recurrent support vector regression model in rainfall forecasting. *Hydrological Process*. 2007; 21: 819-827.
- [11] Kihani A, Akhundali A, Fathian H. Uncertainty analysis of parameters of SVM model for estimation of suspended sediment load and bed in Sierra Karaj station with Monte Carlo simulation method. *Iran Water and Soil Research (Agricultural Sciences of Iran)*. 2021; 52(1): 195-212. [Persian].
- [12] Breiman L. Application and analysis of random forests and machine learning. *Journal of Water Management*. 2001; 15(1): 5-32
- [13] Pal M. M5 model tree for land cover classification. *International Journal of Remote Sensing*, 2006; 27(4): 825-831.
- [14] Javidan S, Sattari MT, Mehrabi A. Comparative analysis of estimating monthly reference evapotranspiration using Kernel and Tree-Based data mining models versus empirical methods. *Environment and Water Engineering*. 2022; 8(4): 908-922.
- [15] Piri H, Mobaraki M. Investigating the effectiveness of transfer functions based on machine learning methods for predicting reference evaporation and transpiration (Case study: Bushehr). *Eco Hydrology*. 2023; 10(3): 421-434. [Persian].
- [16] Seifzadeh KH, Zare Haghi D, Samadianfard S, Neyshabouri MR, Mikaeili F. Estimation of daily evaporation from class A pan using five data mining methods (case study: tabriz synoptic station. *Water and Soil Science*. 2023; 33(4): 167-184. [Persian].
- [17] Sattari MT, Nahrin F, Azimi V. Prediction of daily reference evapotranspiration using artificial neural network model and M5 tree model (Case study: Bonab Station). *Iranian Journal of Irrigation and Drainage*. 2013; 1(7): 104-113. [Persian].
- [18] Siaser H, Dindarlou A. Estimation of daily reference evaporation and transpiration using deep learning model, random forest and decision tree (case study: Sistan plain). *Iranian Water Research*, 2019; 14(1): 108-99.

- [19] Izadyar M, Samadianfard S, Majnooni Heris A, Sadraddini AA. Estimating Pan Evaporation in Three Different Climates using Data-Driven. *Water and Soil Science*, 2024; 34(1): 113-132. [Persian].
- [20] *Integrated Management Plan for Iran's Coastal Areas (Process and Achievements)*, Deputy Director of Port Development and Equipment, General Directorate of Coastal and Port Engineering. 2013. [Persian].
- [21] Tabari H., Grismer M. E., Trajkovic S. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science*. 2013; 31: 107-117.
- [22] Allen RG, Pereira LS, Raes D, Smith M. Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements. *Irrigation and Drainage Paper* No. 56, FAO, Rome, Italy, 1998; 300 pp.
- [23] Farshadfar A. *Multivariate statistical principles and methods*. Publications of Razi University of Kermanshah. 2014. [Persian].
- [24] Azimi A, Rangzan K, Kaboulizade M, Khoramian M. Estimating of evapotranspiration using remote sensing, artificial neural network and comparison with the experimental method (Penman-Monteith-FAO). *Journal of RS and GIS for Natural Resources*. (2016); 6(4): 61-75.
- [25] Burges C.J. A. tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*. 1998; 2: 121-167
- [26] Norouzi Ghoshbalagh H, Nadiri A, Asghari Moghaddam A, Qarahkhani M. Comparison of the efficiency of artificial neural networks, fuzzy logic and random forest in estimating the aquifer transfer capability of Malekan plain. *Echo Hydrology*. 2018; 5(3): 739-751. [Persian]
- [27] Nosrati Karizak F, Movahedi Naeni SA, Hezarjaribi A. Using Artificial Neural Networks to estimate saturated hydraulic conductivity from easily available soil properties. *Journal of Soil Management and Sustainable Production*. 2012; 2(1): 95-110. [Persian].
- [28] Wosten JHM, Pachepsky Ya A, Rawls WJ. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*. 2001; 251: 123-150.
- [29] Debeljak M, Džeroski S. Decision Trees in Ecological Modelling. In: Jopp, F., Reuter, H., Breckling, B. (eds), *Modelling Complex Ecological Dynamics*. 2011; (197-209). Springer, Berlin, Heidelberg.
- [30] Wang Y, Witten I.H. Inducing model trees for continuous classes. *In Proceedings of the Ninth European Conference on Machine Learning*. 1997; 128-137.
- [31] Breiman L. Application and analysis of random forests and machine learning. *Journal of Water Management*. 2001; 15(1): 5-32
- [32] Siasar H, Honar T. The application of support vector machine, chaid and random forest models in estimating daily reference transpiration evaporation in the north of Sistan and Baluchistan province. *Iran Irrigation and Drainage*. 2018; 2(13): 378-388. [Persian].
- [33] Joudy AR, Sattari MT. Estimation of scour depth of piers in hydraulic structures using Gaussian process regression. *Irrigation and Drainage Structures Engineering Research* 2015; 16(65): 19-36. [Persian].
- [34] Kuss M. *Gaussian process models for robust regression, classification, and reinforcement learning*. Ph. D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany. 2006.
- [35] Saghebian M, Roushangar K. Prediction of total and bedform roughness coefficient in alluvial channels based on experimental data via Gaussian process regression method. *Iranian Journal of Irrigation and Drainage*. 2019; 2: 437-499. [Persian].
- [36] Mouneskhah V., Samadian Fard S., Mouneskhah M.H. Evaluation of Data Mining Methods and Experimental Temperature-Radiation-Based Models in Estimating Evaporation from the Pan (Case Study: East of Urmia Lake). *Iran water and soil research*. 2020; 51(9): 2348-2337.
- [37] Seifzadeh KH, Zare Haghi D, Samadianfard S, Neyshabouri MR, Mikaeili F. Estimation of Daily Evaporation from Class A Pan Using Five Data Mining Methods (Case Study: Tabriz Synoptic Station). *Water and Soil Science*. 2023; 3(4): 167-184.
- [38] Bidabadi M, Babazadeh H, Shiri J, Saremi A. Estimation of reference crop evapotranspiration (ET<sub>0</sub>) using artificial intelligence model in an arid climate with external data. *Applied Water Science*. 2024; 14: 3.

- [39] Soltani A, Mirltifi M, Dehghani Sanij H. Estimation of reference evapotranspiration using limited meteorological data in different climatic journal. *Water and soil*. 2013; 26(1): 139-149
- [40] Mousavi Baighi M, Erfanian M, Sarmad M. Using minimum meteorological data to estimate reference plant evapotranspiration and provide correction coefficient (case study: Razavi Khorasan province). *Water and Soil Journal*. 2018; 23(1): 91-99.