

Evaluation of Data Mining Methods and Experimental Temperature-Radiation-Based Models in Estimating Evaporation from the Pan (Case Study: East of Urmia Lake)

VAHID MOUNESKHAH¹, SAEID SAMADIANFARD¹, MOEIN HADI^{1*}

1. Department of Water Engineering, Tabriz University, Tabriz, Iran.

(Received: May. 24, 2020- Revised: July. 9, 2020- Accepted: July. 14, 2020)

ABSTRACT

Evaporation from the pan has an effective role in water resources management. But due to the interaction of meteorological variables in the calculation of evaporation, several nonlinear relationships have been presented that their efficiency is arguable according to the climatic conditions of each region. Therefore, in the present study, the capabilities of temperature-radiation-based empirical equations and data mining methods of support vector regression (SVR), Gaussian process regression (GPR) and nearest neighborhood (IBK) were investigated under 10 different scenarios resulting from the combination of meteorological factors in estimating and predicting the evaporation amounts in 5 selected stations in the east of Urmia Lake basin. NRMSE and MAPE statistical indicators were used to evaluate the results. In order to model the effective parameters on pan evaporation, the effect of each parameter was calculated using the principal component analysis through the correlation values of parameters with the pan evaporation rate. The results showed that among the implemented meteorological parameters, temperature have the maximum impact and wind speed and precipitation have the minimum impacts on modeling process. Also, among the empirical methods, the Jensen-Haise method had the highest accuracy. Moreover, among the data mining methods, the SVR in Tabriz, Sarab, and Harris stations and GPR in Bostanabad and Maragheh stations had higher accuracies as compared to the others. In general, in all the studied stations, the accuracy of the best data mining scenario was higher than the best empirical method. Also, in terms of data limitation, the Jensen-Haise method had suitable accuracy. Also, despite the low accuracy of the IBK method compared to other data mining methods, this method reaches to its highest accuracy rates with the lowest input variable.

Keywords: Data Mining, Jensen-Haise, Modeling, Pan Evaporation, Temperature.

ارزیابی روش‌های داده‌کاوی و مدل‌های تجربی مبتنی بر دما-تشنش در برآورد تبخیر از تشت (مطالعه موردی: شرق دریاچه ارومیه)

وحید مونس‌خواه^۱، سعید صمدیان‌فرد^۱، معین هادی^{*۱}

۱. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

(تایخ دریافت: ۱۳۹۹/۲/۱۵ - تاریخ بازنگری: ۱۳۹۹/۴/۱۹ - تاریخ تصویب: ۱۳۹۹/۴/۲۴)

چکیده

تبخیر از تشت نقش مؤثری در مدیریت منابع آب دارد. ولی به دلیل اثرات متقابل متغیرهای هواشناسی در محاسبه تبخیر، روابط غیرخطی متعددی ارائه شده است که با توجه به شرایط اقلیمی هر منطقه کارایی آن‌ها قابل بحث است. لذا در مطالعه حاضر، کارایی روش‌های تجربی مبتنی بر دما-تشنش و روش‌های داده‌کاوی رگرسیون ماشین بردار پشتیبان (SVR)، رگرسیون فرآیند گاوسی (GPR) و نزدیکترین همسایگی (IBK) تحت ۱۰ سناریو مختلف حاصل از ترکیب عوامل هواشناسی در پیش‌بینی و مدل‌سازی تبخیر از تشت در ۵ ایستگاه منتخب در شرق حوضه دریاچه ارومیه بررسی شد. برای ارزیابی نتایج از شاخص‌های آماری NRMSE و MAPE استفاده شد. به منظور مدل‌سازی پارامترهای مؤثر در تبخیر از تشت، میزان تأثیر هر یک از پارامترها با استفاده از روش تجزیه به مؤلفه‌های اصلی از طریق مقادیر همبستگی پارامترها با میزان تبخیر از تشت محاسبه گردید. نتایج نشان داد در بین متغیرهای هواشناسی مورد بررسی، دما بیشترین و سرعت باد و بارش کمترین تأثیر را در مدل‌سازی دارند. همچنین در بین روش‌های تجربی، روش جنسن-هیز دارای بالاترین دقت بود. علاوه بر این، در بین روش‌های داده‌کاوی نیز روش SVR در ایستگاه‌های تبریز، سراب و هریس و روش GPR در ایستگاه‌های بستان‌آباد و مراغه در مقایسه با سایر روش‌ها دقت بالاتری داشتند. به طور کلی در تمام ایستگاه‌ها دقت بهترین سناریوی روش‌های داده‌کاوی بالاتر از بهترین روش تجربی بود. در شرایط محدودیت داده نیز روش جنسن-هیز دقت مطلوبی داشت. همچنین، علی‌رغم دقت پایین روش IBK نسبت به سایر روش‌های داده‌کاوی، این روش با متغیرهای ورودی کمتری به بالاترین دقت خود در مدل‌سازی تبخیر می‌رسد.

واژه‌های کلیدی: تبخیر از تشت، جنسن-هیز، داده‌کاوی، دما، مدل‌سازی.

مقدمه

تبخیر از تشت به عنوان یکی از مؤلفه‌های اصلی چرخه آب، نقش اساسی در مطالعات کشاورزی، هیدرولوژی، برنامه‌ریزی آبیاری و مدیریت منابع آب ایفا می‌کند (Gundekar et al., 2008). همچنین، پدیده تبخیر موجب افزایش غلظت نمک و کاهش کیفیت آب می‌گردد. تبخیر به عنوان یک فرآیند فیزیکی، رابطه‌ای قوی با عوامل جوی دارد که از مهم‌ترین آن‌ها می‌توان به دما، تابش خورشیدی، رطوبت نسبی و باد اشاره کرد (Chow et al., 1988). پژوهشگران توانسته‌اند با استفاده از روش‌های ریاضی، تجربی و ترکیب آن‌ها، به تحلیل پدیده تبخیر بپردازند (Coulomb et al., 2001; Cohen et al., 2002; Gavin and Agnew, 2004). با توجه به اهمیت تبخیر در چرخه آب و تأثیر آن بر کمیت و کیفیت منابع آب‌های سطحی، مطالعه و شناخت دقیق این پدیده به عنوان یکی از مسائل مهم در مطالعات منابع آبی مطرح است. با استفاده

از تشت تبخیر به عنوان وسیله‌ای کم‌هزینه، میزان تبخیر از سطح آب به صورت مستقیم قابل اندازه‌گیری است. اما با توجه به این که شمار ایستگاه‌های تبخیرسنجی کافی نبوده و توزیع مکانی مناسبی هم ندارند، روش‌های جایگزین یا روش‌های غیرمستقیم از جمله استفاده از روابط تجربی، الگوریتم‌های هوش مصنوعی و تکنیک سنجش از دور به منظور برآورد میزان تبخیر از سطوح آزاد آبی مورد توجه بسیاری از محققان قرار گرفته است (Rosenberry et al., 2007; Sun et al., 2011; Hassan, 2013; Dalkilic et al., 2014; Akbarzadeh et al., 2015). نتایج پژوهشی در منطقه کرمان نشان داد که روش فازی-عصبی نسبت به بقیه روش‌ها از دقت بیشتری برای برآورد تبخیر از تشت برخوردار است (Shadmani and Marofi., 2011). در مطالعه‌ای دیگر با استفاده از داده‌های تشت تبخیر و سری زمانی، روند تغییرات فصلی تبخیر

کلاسه‌بندی‌های آماری به کار گرفته شده است (Sharifazari, 1392). Singh *et al.* (2005) از روش نزدیک‌ترین همسایه برای پیش‌بینی وضعیت آب و هوایی در شمال هند استفاده کرده و دقت بالای این روش را گزارش کردند. در مطالعه‌های دیگر، ارزیابی هفت روش برآورد تبخیر نشان داد که مدل‌های ساده که صرفاً متکی بر داده‌های تابش خورشیدی و دمای هوا هستند، نتایج بهتری را در مقایسه با روش‌های ترکیبی که به داده‌های متعددی نیاز دارند حاصل می‌کند (Abtew, 2001). مطالعه کارآیی تکنیک‌های داده‌محور از جمله رگرسیون بردار پشتیبان (SVR) و شبکه‌های عصبی مصنوعی (ANN) و ترکیب آن‌ها با تبدیل موجک (WSVR و WANN) برای پیش‌بینی نرخ تبخیر نشان داد، تبدیل موجک در افزایش دقت پیش‌بینی‌های ANN و SVR اثرات مثبتی نداشته و شبکه عصبی روش مناسبی برای مدل‌سازی تبخیر در ایستگاه تبریز است (Qasem *et al.*, 2019). در پژوهشی دیگر، مطالعه روابط تجربی برآورد تبخیر، نشان داد که روش جنسن-هیز بهترین روش برآورد تبخیر از سطوح آزاد آبی در حوضه دریاچه ارومیه است (Mouneskhah *et al.*, 2018).

با توجه به اهمیت پدیده تبخیر در مطالعات هیدرولوژیکی و مدیریت بهینه مصرف آب، در مطالعه حاضر کارآیی روش‌های تجربی بر پایه دما-تشنع در کنار روش‌های داده‌کاوی در برآورد مقادیر تبخیر از تشت و مدل‌سازی آن با توجه به عوامل مؤثر بر آن مورد ارزیابی قرار گرفتند.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

تحقیق حاضر با استفاده از داده‌های هواشناسی ایستگاه‌های سینوپتیک منتخب در شرق حوضه دریاچه ارومیه شامل تبریز، بستان‌آباد، هریس، مراغه و سراب انجام گرفت. ملاک انتخاب ایستگاه‌ها وجود داده‌های پیوسته و بلندمدت است. در جدول (۱) مشخصات ایستگاه‌های سینوپتیک و طول دوره داده‌برداری در هر ایستگاه (با توجه به محدودیت داده) ارائه شده است.

بررسی شد. نتایج نشان داد به دلیل محدودیت‌های داده و فرضیه‌های به کار گرفته‌شده در روش سری زمانی، مدل‌های مورد استفاده برای پیش‌بینی‌های کوتاه‌مدت قابل اطمینان هستند (Bahmani *et al.*, 2013). Ghahreman and Gharekhani (2011) نیز با ارزیابی توانمندی مدل‌های سری زمانی ARIMA در برآورد تبخیر از تشت گزارش کردند مدل سری زمانی ARIMA(1,1,1) نتایج بسیار بهتری نسبت به سایر مدل‌های ARIMA داشته و روند تغییرات سری زمانی را بهتر شبیه‌سازی می‌کند. همچنین نتایج ایشان نشان داد که هیچ‌یک از مدل‌های سری زمانی برای پیش‌بینی‌های بلندمدت توصیه نمی‌شود. Terzi (2011) با استفاده از روش برنامه‌ریزی ژنتیک، تبخیر روزانه از سطح دریاچه اغیردیر ترکیه را مدل کرده و گزارش نمود که مدل برنامه‌ریزی ژنتیک می‌تواند جایگزین روش تشت تبخیر گردد. Tabari *et al.* (2010) مدل‌های رگرسیون چندگانه و شبکه عصبی مصنوعی را برای برآورد تبخیر از تشت در استان همدان مورد بررسی قرار داده و گزارش کردند مدل‌های شبکه عصبی در مقایسه با رگرسیون چندگانه برتری نسبی دارند. در مطالعه‌های دیگر، میزان تبخیر از مخازن آب شیرین چاه‌نیمه‌ها با الگوریتم سیال با استفاده از تکنیک سنجش از دور برآورد گردید و نتایج نشان داد که روش سیال ابزار قابل قبولی برای برآورد تبخیر از سطوح آزاد آب است و نتایج به دست آمده می‌تواند مبنای تصمیمات برنامه‌ریزی و مدیریت یکپارچه منطقه قرار گیرد (Akbarzadeh *et al.*, 2015). Samadianfard *et al.* (2018) نیز از روش‌های یادگیری ماشینی برنامه‌ریزی ژنتیک، رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی برای پیش‌بینی مقادیر روزانه تبخیر از تشت در دو ایستگاه هواشناسی تبریز و جلفا استفاده کردند. نتایج نشان داد که در شهر تبریز سناریو دوم روش ANN و در شهر جلفا سناریو هفتم روش ANN کمترین خطا را در مقایسه با سایر سناریوهای تعریف شده دارند و دقت به مراتب بالاتری را دارا می‌باشند. در بین روش‌های داده‌کاوی، اخیراً روشی موسوم به KNN یا نزدیک‌ترین همسایگی نیز برای تولید داده‌های آب و هوایی استفاده شده است که جهت تشخیص الگو و

جدول ۱- مشخصات ایستگاه‌های سینوپتیک

ایستگاه	طول جغرافیایی	عرض جغرافیایی	ارتفاع از سطح آزاد دریا (متر)	دوره داده‌برداری
سراب	۴۷° ۳۲' E	۳۷° ۵۶' N	۱۶۸۲	۱۳۷۵-۱۳۹۷
تبریز	۴۶° ۱۷' E	۳۸° ۰۵' N	۱۳۶۱	۱۳۷۴-۱۳۹۷
بستان‌آباد	۴۶° ۵۱' E	۳۷° ۵۱' N	۱۷۵۰	۱۳۸۷-۱۳۹۷
مراغه	۴۶° ۱۶' E	۳۷° ۲۴' N	۱۴۱۲	۱۳۷۴-۱۳۹۷
هریس	۴۷° ۰۶' E	۳۸° ۱۵' N	۱۸۱۵	۱۳۸۹-۱۳۹۷

از روش‌ها در مقیاس ماهانه، در ۶ ماه سال از اردیبهشت تا مهر که داده‌های تشت تبخیر ثبت شده است، محاسبه گردید و معیار ارزیابی قرار گرفت. لازم به ذکر است با توجه به شرایط اقلیمی منطقه، مقدار تبخیر در ۶ ماهی که داده وجود ندارد کمتر است و به دلیل احتمال وقوع یخبندان به خصوص در فصل زمستان اغلب تشت‌های تبخیر جمع‌آوری می‌شوند و امکان داده‌برداری وجود ندارد.

روابط تجربی مبتنی بر دما-تشنع

روابط تجربی متعددی در نواحی مختلف جهان به منظور برآورد تبخیر ارائه شده است. در مطالعه حاضر از روش‌های دسته دما-تشنع استفاده شد که معادلات و اسامی آن‌ها در جدول (۲) ارائه گردیده است. این گروه از روابط که اغلب مبنای تجربی دارند، در عین سادگی و پارامترهای ورودی موردنیاز کم، با توجه به شرایط محل مورد مطالعه دقت متفاوتی دارند.

با توجه به بحران روزافزون آب در منطقه شمال غرب، در پژوهش حاضر برآورد و پیش‌بینی میزان تبخیر مورد توجه قرار گرفته است. برای این منظور، با توجه به کمبود ایستگاه‌های تبخیرسنجی و عدم وجود داده کافی، از روش‌های داده‌کاوی و روابط تجربی مبتنی بر دما-تشنع استفاده شده و نتایج با داده‌های تشت تبخیر مقایسه گردید. داده‌های مورد استفاده برای انجام پژوهش شامل دمای کمینه، دمای بیشینه، دمای متوسط، ساعات آفتابی، رطوبت نسبی متوسط، رطوبت نسبی کمینه، رطوبت نسبی بیشینه، سرعت باد، بارش، تابش خورشیدی و تبخیر از تشت از سازمان هواشناسی کشور اخذ گردید. داده‌ها بعد از صحت‌سنجی و تکمیل نقاط مفقود، برای مدل‌سازی تبخیر مورد استفاده قرار گرفتند. به منظور تأثیر تمام داده‌ها در آموزش و صحت‌سنجی مدل‌های توسعه داده شده، مقادیر ۷۵ و ۲۵ درصد داده‌ها به ترتیب برای آموزش و صحت‌سنجی در نظر گرفته شد. به منظور همسان‌سازی نتایج و امکان مقایسه، میزان تبخیر حاصل

جدول ۲- معادلات و اسامی روش‌های تجربی مبتنی بر دما-تشنع

معادله	منبع	نام روش
$E = (0.014 T_a - 0.37)(Q_s \times 3.523 \times 10^{-2})$	McGuinness and Bordne (1972)	جنسن-هیز
$E = \left(52.6 \frac{S}{S + \gamma} \frac{Q_s}{\gamma \cdot L \cdot \rho} \right) - 0.12$	McGuinness and Bordne (1972)	مکینک
$E = (0.0082 T_a - 0.19)(Q_s \times 3.495 \times 10^{-2})$	McGuinness and Bordne (1972)	استیفنز-استوارت

پشتیبان در حل مسائل پیش‌بینی کاربرد دارد (Fallahi et al., 2012). در یک مدل رگرسیونی SVR لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز مشخص می‌شود.

$$y = f(x) + \text{noise} \quad (\text{رابطه ۱})$$

بنابراین، موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح موارد جدیدی را که SVR تاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به وسیله آموزش مدل SVR بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی به منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVR شناخته شده است که عبارتند از: مدل‌های رگرسیونی SVR نوع اول که به مدل‌های SVR -v مشهورند و مدل‌های رگرسیونی SVR نوع دوم که به مدل‌های SVR -ε مشهورند. در این مطالعه، مدل SVR -

در این روابط E مقدار تبخیر ماهانه (میلی‌متر)، T_a دمای هوا (درجه فارنهایت)، Q_s تابش خورشیدی (وات بر مترمربع)، S شیب منحنی فشار بخار اشباع در دمای متوسط هوا (پاسکال بر درجه سانتی‌گراد)، γ ثابت سایکرومتری (پاسکال بر درجه سانتی-گراد)، L گرمای نهان تبخیر (مگا ژول بر کیلوگرم)، ρ چگالی آب (کیلوگرم بر مترمکعب) می‌باشند.

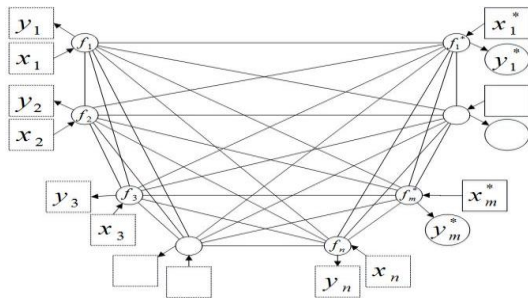
رگرسیون ماشین بردار پشتیبان (SVR)

ماشین بردار پشتیبان یکی از روش‌های یادگیری است که بر مبنای تئوری یادگیری آماری در سال ۱۹۹۲ میلادی معرفی شده است (Boser et al., 1992). گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ به نتیجه رسید (Vapnik, 1995). ماشین بردار پشتیبان مبتنی بر کمینه کردن ساختاری ریسک است که از نظریه آموزش آماری گرفته شده است (Vapnik, 1998). مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده مدل طبقه‌بندی ماشین بردار پشتیبان و مدل رگرسیون بردار پشتیبان تقسیم‌بندی می‌شوند. مدل رگرسیون بردار

داشته به نحوی که $p(f|X)=N(0,K)$ که در آن K ماتریس کواریانس با درایه‌های $k_{i,j}=k(x_i,x_j)$ است.

$$K(X,X)=\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{pmatrix} \quad (\text{رابطه ۸})$$

مقدار $k_{i,j}$ کواریانس بین مقادیر توابع نهان $f(x_i)$ و $f(x_j)$ است. رگرسیون فرآیند گاوسی به منظور محاسبه توزیع پیش‌بینی شده برای مقادیر تابع f^* در نقاط تست $X^*=[x_1^*,x_2^*,\dots,x_m^*]$ به کار می‌رود. مدل تصویری فرآیند گاوسی در شکل (۱) ارائه شده است. در این شکل f_i بیانگر $f(x_i)$ است. مجموعه توابع نهان f_i که با شاخص x_i نشان داده شده‌اند، به‌طور کامل به یکدیگر مرتبط می‌باشند. هر اتصال نشان‌دهنده یک رابطه بین دو متغیر نهان بوده که توسط تابع کواریانس تعریف می‌شود.



شکل ۱- مدل تصویری رگرسیون فرآیند گاوسی

در شکل (۱) مربعات نشان‌دهنده متغیرهای مشاهداتی و دایره‌های نمایانگر مجهولات می‌باشند. توزیع y مشروط به مقادیر f بوده که با یک گاوسین ایزوتروپیک به شکل زیر ارائه می‌شود:

$$p(y|f,X)=N(f,\sigma_n^2 I) \quad (\text{رابطه ۹})$$

که در آن: I ماتریس همانی است. با توجه به ویژگی‌های تابع گاوسی، توزیع حاشیه‌ای y به شکل زیر تعیین می‌شود:

$$p(y|X)=\int p(y|f,X)p(f|X)df=N(0,K+\sigma_n^2 I) \quad (\text{رابطه ۱۰})$$

توزیع ادغام‌شده مقادیر مشاهداتی که خروجی موردنظر می‌باشند و مقادیر تابع در نقاط تست به‌صورت زیر نوشته می‌شوند:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X)+\sigma^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \quad (\text{رابطه ۱۱})$$

با توجه به رابطه (۶) و استفاده از قواعد استاندارد برای مقید نمودن گاوسی، می‌توان توزیع شرطی را بدست آورد:

$$p(f_*|X,y,X_*) \sim N(\bar{f}_*, \text{cov}(f_*)) \quad (\text{رابطه ۱۲})$$

$$\bar{f}_* = K(X_*,X)[K(X,X)+\sigma^2 I]^{-1}y \quad (\text{رابطه ۱۳})$$

$$\text{cov}(f_*) = K(X_*,X_*) - K(X_*,X)[K(X,X)+\sigma^2 I]^{-1}K(X,X_*) \quad (\text{رابطه ۱۴})$$

ε به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده گردید. برای این مدل، تابع خطا به‌صورت زیر تعریف می‌شود.

$$y_i - W^T \phi(X_i) + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (\text{رابطه ۲})$$

تابع خطای مذکور لازم است که با توجه به محدودیت‌های زیر کمینه گردد.

$$W^T \phi(X_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (\text{رابطه ۳})$$

$$y_i - W^T \phi(X_i) - b \leq \varepsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0, \quad i=1, \dots, N$$

که در آن: C ثابت گنجایش، W بردار ضرایب، W^T ترانزپوز بردار ضرایب، ξ_i و ξ_i^* ضرایب کمبود، b ضریبی ثابت، N الگوی آموزش مدل و ϕ تابع کرنل است (Vapnik, 1998). از بین همه توابع کرنل موجود در این روش، بهترین تابع بر حسب کمترین خطا انتخاب شد.

رگرسیون فرآیند گاوسی (GPR)

مجموعه داده S با n مشاهده را در نظر بگیرید $S = \{(x_i, y_i) | i=1, \dots, n\}$ که در آن بردار ورودی با D بعد و y_i خروجی اسکالر یا هدف است. این مجموعه متشکل از دو جز ورودی و خروجی به‌عنوان نقاط نمونه یا تجربی معرفی خواهند شد. به‌منظور سهولت کار، ورودی‌های مجموعه در ماتریس $X=[x_1, x_2, \dots, x_n]$ و خروجی‌ها در ماتریس $Y=[y_1, y_2, \dots, y_n]$ جمع می‌گردند. وظیفه رگرسیون، ایجاد یک ورودی جدید x^* به‌منظور دستیابی به توزیع پیش‌بینی شده برای مقادیر متناظر داده‌های مشاهداتی y^* و بر مبنای مجموعه داده S است. فرآیند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد محدودی از آن‌ها با توزیع‌های گاوسی ادغام شده‌اند. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی است؛ اما فرآیند گاوسی بیانگر توزیع بین توابع است. فرآیند گاوسی $f(x)$ توسط توابع میانگین $m(x)$ و کواریانس به شکل زیر تعریف می‌شود (Kuss, 2006):

$$m(x) = E(f(x)) \quad (\text{رابطه ۴})$$

$$k(x, x') = E((f(x) - m(x))(f(x') - m(x'))) \quad (\text{رابطه ۵})$$

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

$$f(x) \sim GP(m(x), k(x, x')) \quad (\text{رابطه ۶})$$

برای ساده‌سازی، مقدار تابع میانگین صفر فرض می‌شود. در فرآیند گاوسی رابطه بین بردار ورودی و هدف به‌فرم زیر است:

$$y_i = f(x_i) + \varepsilon \quad (\text{رابطه ۷})$$

که در آن $f(x)$ بیانگر تابع رگرسیون دلخواه و ε نیز مقدار نویز توزیع گاوسی با میانگین صفر و واریانس σ^2 است، یعنی $\varepsilon \sim N(0, \sigma^2)$. علاوه بر این، چنین فرض می‌گردد که

$$f = [f(x_1), f(x_2), \dots, f(x_n)]^T$$

تجربی، از شاخص‌های آماری ضریب همبستگی (r)، جذر ریشه میانگین مربع خطا نرمال شده (NRMSE) و میانگین قدر مطلق درصد خطا (MAPE) برای تعیین میزان همبستگی بین مقادیر تشت تبخیر و مقادیر برآورد شده از مدل‌ها و تعیین مقادیر خطای برآورد شده استفاده شد:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{رابطه ۱۶})$$

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}}{\bar{y}} \times 100 \quad (\text{رابطه ۱۷})$$

$$\text{MAPE} = \frac{\sum_{i=1}^n \left(\frac{|x_i - y_i|}{x_i}\right)}{n} \times 100 \quad (\text{رابطه ۱۸})$$

در روابط ۱۶ تا ۱۸، x_i و y_i به ترتیب i امین داده مشاهداتی و برآورد شده، \bar{x} و \bar{y} به ترتیب میانگین داده‌های مشاهداتی و برآورد شده و n تعداد گام‌های زمانی هستند. لازم به ذکر است که معیار NRMSE زیر ۱۰ درصد نشان‌دهنده دقیق بودن مدل، ۲۰-۳۰ درصد مناسب بودن مدل، ۳۰-۴۰ درصد نشانه ضعیف بودن مدل است (Singh et al., 2008).

نتایج و بحث

کارایی روش‌های تجربی و داده‌کاوی در برآورد میزان تبخیر از تشت در شرق حوضه دریاچه ارومیه مورد ارزیابی قرار گرفت. برای این منظور مقادیر ۷۵ و ۲۵ درصد داده‌ها به ترتیب برای آموزش و صحت‌سنجی استفاده شد. مقادیر تبخیر از تشت با استفاده از روش‌های تجربی مبتنی بر دما-تشنش محاسبه شده و با داده‌های تشت تبخیر مقایسه گردید. با توجه به شاخص‌های خطای مورد استفاده، نتایج در جدول (۳) ارائه شده است.

در همه ایستگاه‌های مورد مطالعه، روش جنسن-هیز با اختلاف چشمگیری نسبت به سایر روش‌ها، بهترین روش تجربی برآورد تبخیر از تشت است. بر مبنای نتایج به‌دست‌آمده از شاخص NRMSE، نتایج به‌دست‌آمده از روش جنسن-هیز در ایستگاه‌های مراغه و بستان‌آباد دقیق (خطا کمتر از ۱۰ درصد) و در تبریز، هریس و سراب مناسب (خطا بین ۱۰ تا ۲۰ درصد) است. مجیدی و همکاران نیز گزارش کردند که در بین ۱۰ روش تجربی بررسی شده با استفاده از برآوردهای بیلان انرژی، روش‌های تابش-دما و روابط ترکیبی، روش جنسن-هیز با RMSD معادل ۱/۲۱ میلی‌متر بر روز رتبه نخست را به خود اختصاص داد (Majidi et al., 2014).

به‌منظور مدل‌سازی پارامترهای مؤثر در تبخیر از تشت،

پایگاه‌های یادگیری با پارامتر k (IBK) یا k -نزدیکترین همسایگی (KNN)

جستجوی نزدیک‌ترین همسایگی (نزدیک‌ترین نقطه) یک مسئله بهینه‌سازی برای پیدا کردن نزدیک‌ترین نقطه‌ها در فضای متریک^۱ است. مجموعه S شامل تعدادی نقطه در یک فضای متریک مانند M و نیز یک نقطه پرس‌وجوی $q \in M$ داده شده است؛ هدف پیدا کردن نزدیک‌ترین نقطه در S به q است. اغلب فضای M به صورت یک فضای اقلیدسی d -بعدی و فاصله بین نقاط با معیار فاصله اقلیدسی سنجیده می‌شود. جستجوی k نزدیک‌ترین همسایه، K همسایه نزدیک‌تر به نقطه پرس‌وجو را برمی‌گرداند. در واقع روش IBK، یک روش تشخیص الگوی آماری بدون متغیر است که برای الگوی هیدرولوژیکی موجود، k الگوی مشابه بنام نزدیک‌تری همسایه‌ها را می‌یابد. این روش در تجزیه و تحلیل پیش‌بینی، به‌منظور تخمین یا دسته‌بندی یک نقطه بر اساس اجماع همسایگان آن استفاده می‌شود. گراف k نزدیک‌ترین همسایه گراف‌بند که در آن هر نقطه در گراف K نزدیک‌ترین همسایگان خود متصل است. به‌طور کلی این الگوریتم برای دو هدف مورد استفاده قرار می‌گیرد؛ برآورد عملکرد چگالی توزیع داده‌های آزمون و طبقه‌بندی داده‌های آزمون بر اساس الگوهای آزمون. اولین قدم برای استفاده از این الگوریتم، یافتن یک روش و رابطه برای محاسبه فاصله بین داده‌های آزمون و آموزش است.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (\text{رابطه ۱۵})$$

متغیر X ، داده‌های آموزش را با پارامترهای مشخص شده (x_n) تا (x_1) و Y داده‌های آموزش را با همان پارامتر (y_n تا y_1) نشان می‌دهد. پس از تعیین فاصله اقلیدسی بین داده‌ها، نمونه‌های پایگاه داده به ترتیب صعودی از حداقل فاصله (حداکثر تشابه) تا حداکثر فاصله (حداقل تشابه) طبقه‌بندی شده‌اند. مرحله بعدی، یافتن تعداد نقاط (k) آزمایش برای برآورد ویژگی‌های پایگاه داده مورد نظر است. تعیین تعداد همسایگان (k) از مهمترین مراحل است و میزان مهارت این روش به انتخاب نزدیک‌ترین (شبیه‌ترین) نمونه‌ها از بانک اطلاعاتی مرجع به میزان قابل‌توجهی بستگی دارد. اگر فرض شود k کوچک باشد، نتایج به نقاط منفرد غیرمتعارف مدل حساس هستند و اگر فرض شود k قابل‌توجه است، می‌توان مقداری از سایر کلاس‌ها را در محدوده مورد نظر قرار داد. معمولاً بهترین مقدار برای k با استفاده از اعتبارسنجی متقاطع محاسبه می‌شود (Wu et al., 2008).

معیارهای ارزیابی نتایج

به‌منظور ارزیابی نتایج به‌دست‌آمده از روش‌های داده‌کاوی و روابط

نشان‌دهنده اثر همسوی تغییرات هرکدام از متغیرها با تبخیر از تشت و مقادیر منفی نیز نشان‌دهنده تأثیر معکوس تغییرات پارامترها بر میزان تبخیر از تشت است. نتایج به‌دست‌آمده از مطالعه مونس خواه و همکاران نیز نشان داد که در بین عوامل هواشناسی مؤثر در تبخیر، دما، ساعات آفتابی و تابش خورشیدی بیشترین تأثیر را دارند (Mouneskhah et al., 2018).

میزان تأثیر هریک از پارامترها با استفاده از روش تجزیه به مؤلفه‌های اصلی از طریق مقادیر همبستگی پارامترها با میزان تبخیر از تشت محاسبه گردید (جدول ۴). نتایج نشان داد متغیرهای دما در تمام ایستگاه‌ها بیشترین همبستگی و دو متغیر سرعت باد و بارش نیز کمترین میزان همبستگی را با مقادیر تبخیر از تشت داشته‌اند. لازم به ذکر است که مقادیر ضرایب مثبت همبستگی

جدول ۳- ارزیابی نتایج به‌دست‌آمده از روش‌های تجربی دما-تشنه

ایستگاه	معیار	مکینک	جنسن-هیز	استیفنز استوارت
تبریز	R	۰/۹۳۰	۰/۹۷۴	۰/۹۷۲
	MAPE (%)	۳۸/۲۹	۱۱/۴۵	۴۴/۶۰
	NRMSE (%)	۴۴/۹۱	۱۳/۷۶	۴۷/۶۱
مراغه	R	۰/۹۲۹	۰/۹۷۰	۰/۹۶۹
	MAPE (%)	۳۲/۸۳	۶/۹۰	۳۹/۴۵
	NRMSE (%)	۳۹/۰۲	۷/۷۳	۴۱/۹۲
سراب	R	۰/۹۴۳	۰/۹۷۳	۰/۹۷۲
	MAPE (%)	۲۹/۴۵	۱۲/۴۲	۴۴/۴۵
	NRMSE (%)	۳۲/۸۸	۱۲/۹۲	۴۵/۱۷
بستان‌آباد	R	۰/۹۰۷	۰/۹۳۳	۰/۹۳۲
	MAPE (%)	۲۴/۸۷	۱۱/۷۷	۳۳/۶۶
	NRMSE (%)	۲۱/۴۹	۹/۷۳	۲۵/۱۳
هریس	R	۰/۹۲۳	۰/۹۴۶	۰/۹۴۵
	MAPE (%)	۲۱/۷۸	۱۱/۰۱	۳۳/۶۴
	NRMSE (%)	۲۸/۲۳	۱۴/۲۵	۳۵/۲۸

جدول ۴- تجزیه به مؤلفه‌های اصلی بر اساس مقادیر همبستگی پارامترها با تبخیر از تشت

ردیف	هریس	بستان‌آباد	سراب	مراغه	تبریز
۱	دمای متوسط ۰/۷۷	دمای متوسط ۰/۷۸	دمای متوسط ۰/۷۵	دمای متوسط ۰/۸۱	دمای کمینه ۰/۸۰
۲	دمای بیشینه ۰/۷۷	دمای بیشینه ۰/۷۶	دمای بیشینه ۰/۷۱	دمای بیشینه ۰/۸۰	دمای متوسط ۰/۷۹
۳	دمای کمینه ۰/۷۳	دمای کمینه ۰/۶۸	دمای کمینه ۰/۶۸	دمای کمینه ۰/۷۹	دمای بیشینه ۰/۷۸
۴	تابش خورشیدی ۰/۶۵	تابش خورشیدی ۰/۶۶	تابش خورشیدی ۰/۶۴	رطوبت نسبی متوسط ۰/۶۸	تابش خورشیدی ۰/۶۶
۵	رطوبت نسبی متوسط ۰/۵۹	رطوبت نسبی متوسط ۰/۶۴	رطوبت نسبی متوسط ۰/۵۲	رطوبت نسبی بیشینه ۰/۶۸	رطوبت نسبی بیشینه ۰/۶۲
۶	رطوبت نسبی کمینه ۰/۵۷	رطوبت نسبی کمینه ۰/۵۷	رطوبت نسبی بیشینه ۰/۵۲	تابش خورشیدی ۰/۶۸	رطوبت نسبی متوسط ۰/۶۲
۷	رطوبت نسبی بیشینه ۰/۴۸	رطوبت نسبی بیشینه ۰/۵۷	ساعات آفتابی ۰/۵۱	رطوبت نسبی کمینه ۰/۵۹	رطوبت نسبی کمینه ۰/۵۳
۸	ساعات آفتابی ۰/۴۸	ساعات آفتابی ۰/۵۴	رطوبت نسبی کمینه ۰/۴۱	ساعات آفتابی ۰/۵۶	ساعات آفتابی ۰/۵۰
۹	سرعت باد ۰/۱۸	بارش ۰/۱۹	سرعت باد ۰/۲۸	سرعت باد ۰/۳۴	سرعت باد ۰/۴۳
۱۰	بارش ۰/۰۸	سرعت باد ۰/۰۹	بارش ۰/۱۹	بارش ۰/۲۵	بارش ۰/۱۸

که بیشترین همبستگی را با مقادیر تبخیر از تشت دارد، استفاده گردید. به‌عنوان نمونه برای ایستگاه تبریز در سناریو چهارم، از داده‌های دمای کمینه، دمای متوسط، دمای بیشینه و تابش خورشیدی برای مدل‌سازی استفاده گردید (جدول ۴). نتایج معیارهای ارزیابی به‌دست‌آمده در ایستگاه‌های مورد مطالعه برای هر سه روش داده‌کاوی به‌ازای ۱۰ سناریوی مدل‌سازی در جدول (۵) ارائه شده است. به‌عنوان نمونه در ایستگاه تبریز مدل‌های GPR8، SVR8 و IBK5 به‌عنوان بهترین سناریوهای مدل‌های مذکور در پیش‌بینی تبخیر از تشت معرفی شدند. به این مفهوم

برای مدل‌سازی مقادیر تبخیر از تشت و بررسی کارایی روش‌های داده‌کاوی شامل GPR، SVR و IBK در برآورد تبخیر از تشت از نتایج جدول (۴) استفاده شد. برای این منظور، معیار-های خطای ارزیابی نتایج برای سه روش مذکور در ۱۰ سناریو محاسبه گردید. بر این اساس در سناریو دهم، تمام متغیرها در مدل‌سازی مورد استفاده قرار گرفتند. در سناریو نهم، اثر متغیر آخر (متغیری که کمترین میزان همبستگی را با مقادیر تبخیر از تشت دارد) در مدل‌سازی حذف گردید. این فرآیند به همین ترتیب ادامه یافت و در نهایت در سناریو اول، تنها از یک متغیر

روش درخت تصمیم نتایج مطلوب‌تری را نسبت به مدل KNN حاصل می‌کند (Khalili Naft Chali et al., 2017). به‌طور کلی، نتایج نشان داد در ایستگاه‌های تبریز، سراب و هریس روش SVR و در ایستگاه‌های بستان‌آباد و مراغه روش GPR در مقایسه با سایر روش‌ها دقت بالاتری دارند. شادمانی و همکاران نیز با تحلیل مقایسه‌ای چهار روش مدل‌سازی تبخیر از تشت در منطقه گلستان شامل روش‌های GPR، IBK، PR و SVR، گزارش کردند روش GPR با دقت بالایی برای برآورد میزان تبخیر از تشت قابل استفاده است (Shabani et al., 2020).

که در روش‌های GPR و SVR با استفاده از داده‌های دمای کمینه، دمای متوسط و دمای بیشینه، تابش خورشیدی، رطوبت نسبی بیشینه، رطوبت نسبی متوسط، رطوبت نسبی کمینه و ساعات آفتابی و در روش IBK با استفاده از ورودی‌های دمای کمینه، دمای متوسط و دمای بیشینه، تابش خورشیدی و رطوبت نسبی بیشینه، مدل دارای بیشترین دقت خود است. نتایج نشان داد که در بین روش‌های داده‌کاوی، روش IBK در مقایسه با سایر روش‌ها دقت کمتری دارد. خلیلی نفت‌چالی و همکاران نیز با مقایسه مدل‌های KNN و درخت تصمیم M5 در پیش‌بینی تبخیر از سطوح آزاد آبی در منطقه بیرجند، نشان داد که در اغلب موارد

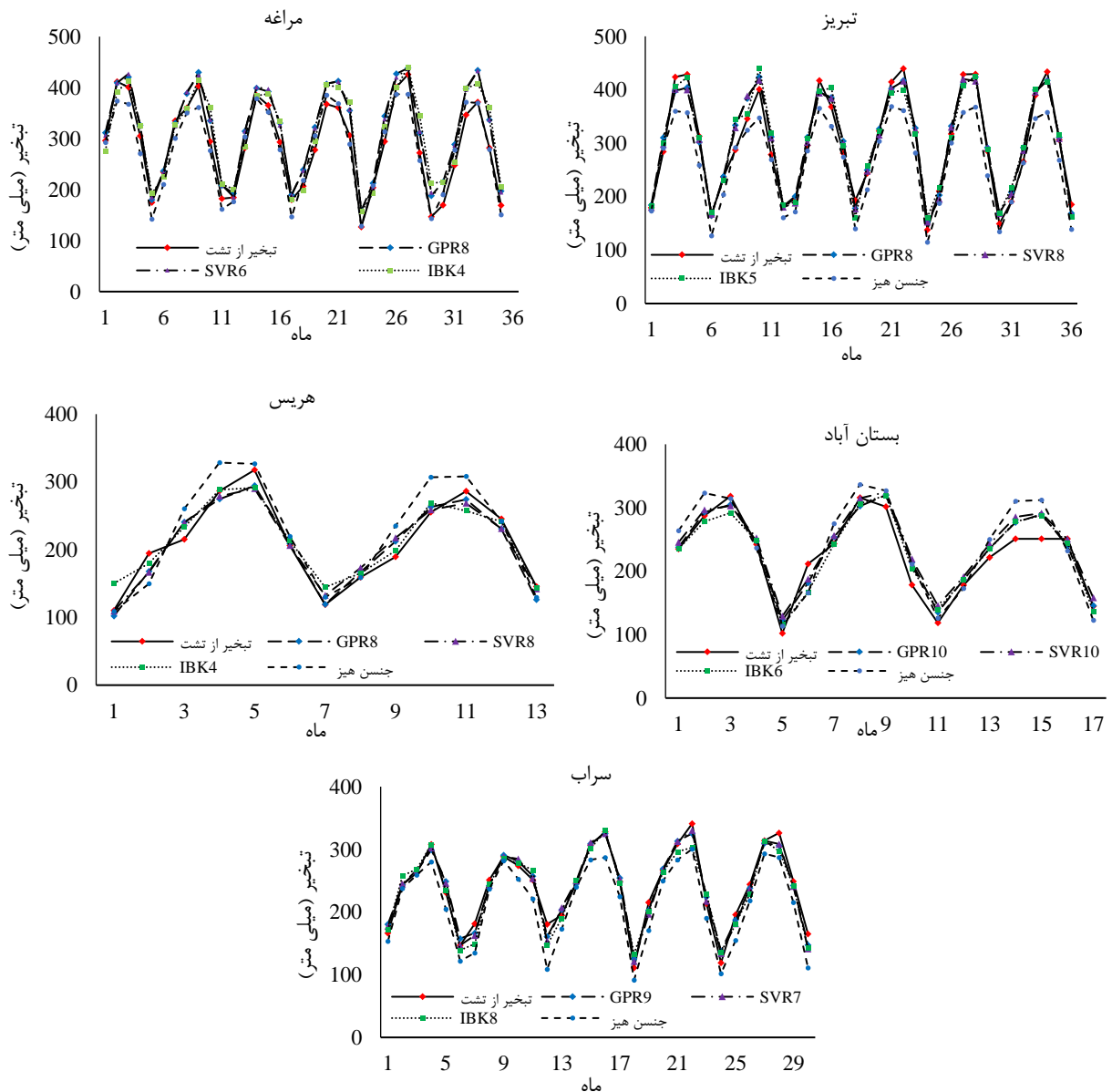
جدول ۵- مقادیر معیارهای ارزیابی روش‌های داده‌کاوی تحت ۱۰ سناریوی مدل‌سازی

ایستگاه	روش	معیار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
GPR		R	۰/۹۵۱	۰/۹۵۶	۰/۹۵۸	۰/۹۷۸	۰/۹۷۸	۰/۹۷۹	۰/۹۷۹	۰/۹۸۱	۰/۹۸۲	۰/۹۸۲
		MAPE (%)	۹/۸۶	۹/۱۳	۸/۹۷	۶/۷۹	۶/۳۲	۶/۱۵	۶/۱۲	۵/۴۲	۵/۸۰	۵/۸۳
		NRMSE (%)	۹/۸۱	۹/۵۵	۹/۳۳	۷/۰۵	۶/۷۸	۶/۵۹	۶/۱۲	۶/۱۵۹	۶/۳۰	۶/۷۷
SVR	تبریز	R	۰/۹۵۰	۰/۹۵۵	۰/۹۵۸	۰/۹۷۶	۰/۹۷۸	۰/۹۸۰	۰/۹۸۰	۰/۹۸۲	۰/۹۸۳	۰/۹۸۳
		MAPE (%)	۱۰/۱۷	۹/۱۲	۸/۹۴	۶/۵۸	۵/۹۲	۵/۶۷	۵/۷۶	۵/۰۹	۷/۳۱	۷/۳۳
		NRMSE (%)	۹/۸۵	۹/۴۰	۹/۰۹	۶/۹۴	۶/۵۳	۶/۲۵	۶/۳۰	۵/۹۱	۷/۲۵	۷/۲۸
IBK		R	۰/۹۴۹	۰/۹۵۴	۰/۹۴۸	۰/۹۶۳	۰/۹۷۵	۰/۹۶۶	۰/۹۵۶	۰/۹۶۹	۰/۹۶۱	۰/۹۶۰
		MAPE (%)	۹/۸۹	۸/۹۲	۱۰/۰۶	۷/۷۵	۶/۱۰	۷/۳۱	۸/۲۸	۷/۳۹	۷/۵۱	۷/۳۸
		NRMSE (%)	۹/۹۴	۹/۶۳	۱۰/۱۲	۸/۶۹	۷/۱۶	۸/۱۳	۹/۳۵	۸/۱۰	۹/۴۵	۹/۴۴
GPR		R	۰/۹۵۵	۰/۹۵۷	۰/۹۵۳	۰/۹۵۴	۰/۹۷۷	۰/۹۷۷	۰/۹۷۷	۰/۹۸۰	۰/۹۸۰	۰/۹۷۹
		MAPE (%)	۱۴/۵۹	۱۳/۹۰	۱۴/۴۲	۱۳/۲۷	۱۳/۴۵	۱۱/۵۵	۱۱/۸۲	۱۱/۳۳	۱۱/۶۳	۱۱/۶۸
		NRMSE (%)	۱۴/۷۱	۱۳/۹۸	۱۴/۵۴	۱۳/۶۹	۱۳/۶۹	۱۳/۶۹	۱۲/۰۶	۱۱/۷۵	۱۲/۶۳	۱۲/۶۶
SVR	مراغه	R	۰/۹۵۴	۰/۹۶۰	۰/۹۵۷	۰/۹۵۹	۰/۹۶۲	۰/۹۷۹	۰/۹۸۰	۰/۹۸۴	۰/۹۸۰	۰/۹۷۹
		MAPE (%)	۱۴/۲۶	۱۳/۲۲	۱۳/۳۷	۱۱/۹۳	۱۱/۵۴	۱۰/۶۳	۱۰/۷۸	۱۰/۶۶	۱۰/۸۶	۱۰/۹۲
		NRMSE (%)	۱۴/۴۴	۱۳/۳۸	۱۳/۶۴	۱۲/۵۳	۱۲/۰۹	۱۱/۴۵	۱۱/۵۹	۱۱/۶۵	۱۲/۴۰	۱۲/۴۴
IBK		R	۰/۹۵۴	۰/۹۵۰	۰/۹۳۹	۰/۹۵۰	۰/۹۴۹	۰/۹۶۴	۰/۹۶۲	۰/۹۶۷	۰/۹۷۱	۰/۹۶۹
		MAPE (%)	۱۴/۴۹	۱۴/۱۴	۱۳/۳۰	۱۰/۸۹	۱۱/۵۵	۱۱/۸۰	۱۱/۱۴	۱۲/۵۵	۱۱/۴۴	۱۰/۹۴
		NRMSE (%)	۱۴/۸۲	۱۴/۵۴	۱۳/۶۴	۱۲/۱۹	۱۳/۳۴	۱۲/۵۳	۱۲/۱۴	۱۳/۰۲	۱۳/۴۶	۱۳/۱۳
GPR		R	۰/۹۳۰	۰/۹۳۸	۰/۹۳۲	۰/۹۵۹	۰/۹۷۸	۰/۹۸۲	۰/۹۸۴	۰/۹۸۳	۰/۹۸۵	۰/۹۸۵
		MAPE (%)	۹/۷۱	۹/۰۲	۹/۳۷	۷/۲۱	۵/۵۰	۵/۱۲	۴/۶۶	۴/۶۳	۴/۵۷	۴/۵۹
		NRMSE (%)	۹/۸۳	۹/۱۵	۹/۴۶	۷/۴۳	۵/۴۱	۵/۰۳	۴/۷۲	۴/۸۶	۴/۵۴	۴/۵۳
SVR	سراب	R	۰/۹۲۸	۰/۹۴۱	۰/۹۳۶	۰/۹۵۸	۰/۹۷۸	۰/۹۷۹	۰/۹۸۴	۰/۹۷۵	۰/۹۷۷	۰/۹۷۷
		MAPE (%)	۹/۶۶	۸/۶۳	۸/۹۳	۷/۱۵	۵/۲۵	۵/۰۵	۴/۳۵	۵/۲۹	۵/۳۲	۵/۲۲
		NRMSE (%)	۹/۸۵۰	۸/۸۱	۹/۰۸	۷/۴۳	۵/۳۲	۵/۲۵	۴/۷۱	۶/۰۳	۶/۰۲	۴/۵۳
IBK		R	۰/۹۲۵	۰/۹۲۴	۰/۹۲۵	۰/۹۲۷	۰/۹۵۵	۰/۹۶۳	۰/۹۶۳	۰/۹۶۹	۰/۹۶۰	۰/۹۶۳
		MAPE (%)	۹/۹۰	۹/۷۵	۹/۸۰	۸/۴۲	۷/۰۱	۷/۱۳	۶/۵۷	۵/۸۹	۶/۹۰	۶/۵۹
		NRMSE (%)	۱۰/۳۱	۹/۹۱	۹/۷۶	۸/۲۳	۷/۶۲	۷/۶۲	۶/۹۶	۶/۶۱	۷/۲۱	۶/۹۵
GPR		R	۰/۹۱۴	۰/۹۱۸	۰/۹۱۲	۰/۹۲۶	۰/۹۳۹	۰/۹۴۱	۰/۹۴۵	۰/۹۵۷	۰/۹۵۸	۰/۹۶۰
		MAPE (%)	۱۱/۶۵	۱۱/۱۹	۱۱/۴۸	۱۰/۰۷	۹/۸۴	۹/۶۹	۹/۹۱	۸/۰۶	۸/۱۱	۷/۸۹
		NRMSE (%)	۸/۳۶	۸/۱۷	۸/۴۱	۷/۶۹	۷/۳۳	۷/۲۷	۷/۱۹	۶/۳۰	۶/۲۳	۵/۹۸
SVR	بستان‌آباد	R	۰/۹۱۸	۰/۹۲۲	۰/۹۱۹	۰/۹۲۸	۰/۹۴۵	۰/۹۴۵	۰/۹۵۳	۰/۹۵۳	۰/۹۶۱	۰/۹۶۱
		MAPE (%)	۱۱/۱۰	۱۰/۶۷	۱۰/۸۹	۱۰/۳۰	۱۰/۲۰	۱۰/۲۴	۱۰/۷۹	۱۰/۰۳	۱۰/۰۵	۹/۹۸
		NRMSE (%)	۸/۰۹	۷/۹۵	۸/۰۸	۷/۷۳	۷/۲۳	۷/۲۵	۷/۴۷	۷/۰۵	۷/۰۶	۷/۰۰
IBK		R	۰/۹۲۵	۰/۹۰۳	۰/۹۱۵	۰/۹۰۲	۰/۹۲۴	۰/۹۴۹	۰/۹۳۶	۰/۹۴۱	۰/۹۴۲	۰/۹۵۰
		MAPE (%)	۱۱/۴۶	۱۱/۸۶	۱۰/۷۷	۱۱/۸۰	۱۰/۱۵	۷/۸۸	۱۰/۰۰	۹/۴۲	۹/۰۹	۸/۰۱
		NRMSE (%)	۷/۹۷	۸/۷۵	۸/۰۲	۸/۶۱	۷/۵۶	۶/۳۰	۷/۳۶	۷/۱۳	۶/۹۲	۶/۳۴
GPR		R	۰/۹۵۰	۰/۹۵۱	۰/۹۴۹	۰/۹۵۰	۰/۹۴۹	۰/۹۵۰	۰/۹۵۲	۰/۹۶۸	۰/۹۶۴	۰/۹۵۸
		MAPE (%)	۱۰/۷۳	۱۰/۳۷	۱۰/۶۷	۹/۵۹	۹/۱۴	۹/۰۹	۸/۸۴	۸/۸۷	۷/۵۰	۷/۸۰
		NRMSE (%)	۱۰/۴۷	۱۰/۲۶	۱۰/۴۹	۹/۹۷	۹/۷۵	۹/۵۸	۹/۴۸	۷/۸۰	۸/۲۵	۸/۲۱
SVR	هریس	R	۰/۹۵۲	۰/۹۵۵	۰/۹۴۹	۰/۹۴۸	۰/۹۵۳	۰/۹۵۹	۰/۹۵۹	۰/۹۶۴	۰/۹۶۱	۰/۹۶۰
		MAPE (%)	۸/۶۶	۸/۲۹	۹/۰۴	۹/۸۱	۹/۱۴	۸/۴۲	۸/۴۳	۸/۸۷	۷/۲۵	۷/۱۰
		NRMSE (%)	۹/۴۷	۹/۱۸	۹/۷۶	۱۰/۲۳	۹/۷۳	۹/۱۶	۹/۲۰	۸/۲۷	۸/۵۶	۸/۵۰
IBK		R	۰/۹۳۶	۰/۹۳۱	۰/۹۵۶	۰/۹۶۷	۰/۹۱۹	۰/۹۳۱	۰/۹۳۱	۰/۹۵۶	۰/۹۴۰	۰/۹۵۷
		MAPE (%)	۱۱/۳۸	۱۲/۰۵	۹/۷۱	۹/۶۶	۱۰/۴۱	۹/۴۲	۹/۴۲	۹/۲۹	۹/۵۹	۹/۶۲
		NRMSE (%)	۱۱/۵۰	۱۲/۲۴	۹/۲۳	۹/۰۵	۱۲/۳۰	۱۱/۲۰	۱۱/۱۹	۱۰/۳۷	۱۱/۰۰	۱۰/۸۵

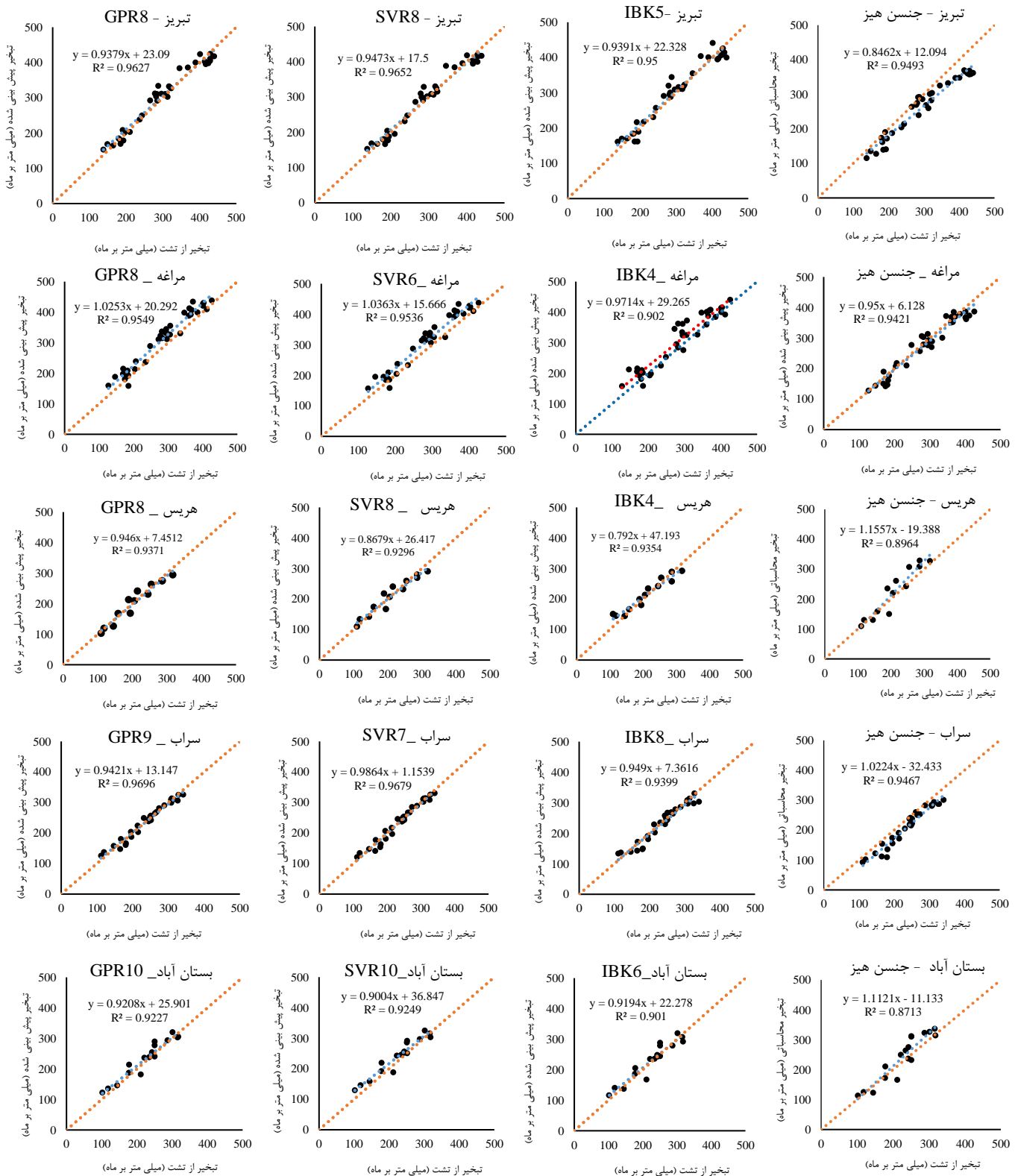
پیش‌بینی‌شده با مقادیر تبخیر از تشت بیشتر بوده و دقت مدل کمتر است.

در اغلب موارد مقایسه نتایج به‌دست‌آمده از روش جنسن-هیز به‌عنوان بهترین روش تجربی دسته دما-تشنع با بهترین سناریوهای روش‌های داده‌کاوی حاکمی از برتری روش‌های داده-کاوی و خطای کمتر آن‌هاست. البته، باید در نظر داشت با توجه به این‌که در اغلب موارد روش‌های داده‌کاوی در سناریوهایی با تعداد داده ورودی بیشتر به مدل به حالت بهینه دست می‌یابند؛ کاربرد روش جنسن-هیز دقت مطلوبی دارد و در شرایط محدودیت داده کاملاً قابل‌اعتناست.

همچنین، روند تغییرات ماهانه مقادیر تبخیر برآورد شده از روش جنسن-هیز به‌عنوان بهترین روش تجربی در تمام ایستگاه‌ها و نیز مقادیر تبخیر پیش‌بینی‌شده از طریق بهترین سناریوی روش‌های GPR، SVR و IBK با مقادیر تبخیر از تشت در شکل (۲) نشان داده شده است. به‌منظور ارزیابی بهتر کارایی روش‌های مورد‌استفاده، نحوه پراکنش مقادیر پیش‌بینی‌شده تبخیر تحت بهترین سناریوی هر مدل حول خط نیمساز و مقایسه آن با مقادیر تبخیر از تشت، در شکل (۳) ارائه شده است. میزان پراکنندگی نقاط پیش‌بینی حاصل از روش‌های مختلف حول خط نیمساز نشان‌دهنده دقت پیش‌بینی هر روش است. به‌طوری‌که هرچه قدر میزان پراکنش از خط نیمساز بیشتر باشد، اختلاف مقادیر



شکل ۲. مقایسه روند تغییرات ماهانه روش جنسن-هیز، بهترین سناریوی روش‌های داده‌کاوی با مقادیر تبخیر از تشت



شکل ۳- پراکنش مقادیر پیش‌بینی شده تبخیر تحت بهترین سناریو هر مدل حول خط نیمساز و مقایسه با مقادیر تبخیر از تشت

نتیجه‌گیری

مبتنی بر دما-تشنوع و روش‌های داده‌کاوی SVR، GPR و

با توجه به اهمیت پدیده تبخیر در مطالعات هیدرولوژیک و مدیریت بهینه منابع آبی، در پژوهش حاضر از روش‌های تجربی

خطا بود. همچنین نتایج حاکی از دقت بالای روش‌های داده-کاوی به خصوص GPR و SVR نسبت به سایر روش‌ها در پیش-بینی مقادیر تبخیر از تشت است. لذا استفاده از این روش‌ها در مدل‌سازی تبخیر از تشت در مناطقی که تعداد ایستگاه‌های تبخیرسنجی کافی نبوده و یا داده‌های ثبت‌نشده زیاد باشد، نتایج قابل‌قبولی خواهد داشت. همچنین کاربرد روش جنسن-هیز در شرایط محدودیت داده نتایج مطلوبی دارد.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

IBK تحت ۱۰ سناریو بر اساس میزان همبستگی و تأثیر متغیر-های ورودی در مدل‌سازی تبخیر از تشت به‌منظور برآورد و پیش‌بینی تبخیر از تشت در ۵ ایستگاه منتخب در شرق حوضه دریاچه ارومیه استفاده شد. در بین عوامل هواشناسی، متغیرهای دمایی بیشترین همبستگی را با تبخیر از تشت داشته و در تمام ایستگاه‌ها بیشترین اولویت را در مدل‌سازی دارند. در بین روابط تجربی دسته دما-تشنش که طبق نتایج بسیاری از پژوهش‌ها از دقیق‌ترین دسته روابط تجربی برآورد تبخیر محسوب می-شوند، روش جنسن‌هیز در تمام ایستگاه‌ها دارای کمترین مقدار

REFERENCES

- Abtew W. (2001). Evaporation estimation for Lake Okeechobee in south Florida. *Journal of Irrigation and Drainage Engineering*, 127, 140-147.
- Akbarzadeh M.S. H., Haghghatjou P. and Bagheri M.H. (2015). Estimates of evaporation from surface water bodies with SEBAL Algorithm using remote sensing techniques (case study: Chahnimeh's Fresh Water Reservoirs of Sistan). *Iranian Journal of Irrigation and Drainage*, 3(9), 510-521. (In Farsi)
- Bahmani R., Radmanesh F., Islamian S.S. and Parham GH. (2013). Reservoir evaporation trend analysis and its prediction using time series. *Journal of Irrigation Sciences and Engineering*, 36(3), 67-80. (In Farsi)
- Boser B.E., Guyon I.M. and Vapnik V.N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler, editor, *5th Annual ACM Workshop on COLT*, Pittsburgh, PA, pp. 144-152.
- Chow V. T., Maidment D. R. and Mays L.W. (1988). *Applied hydrology*. McGraw hill, Newyork, 570 p.
- Cohen S., Ianetz A. and Stanhill G. (2002). Evaporative climate changes at bet Dagon, Israel, 1964-1998, *Agricultural and Forest Meteorology*, 111, 83-91.
- Coulomb C.V., legesse D., Gasse F., Travi Y. and Chernet T. (2001). Lake evaporation estimates in tropical Africal (Lake Ziway, Ethiopia). *Journal of Hydrology*, 245, 1-18.
- Dalkilic Y, Okkan U and Baykan N. (2014). Comparison of different ANN approaches in daily pan evaporation prediction. *Journal of Water Resource and Protection*, 6(4), 319-326.
- Fallahi M.R., Varvani H. and Goliyan S. (2012). Precipitation forecasting using regression tree model to flood control. *5th national conference on watershed & soil and water management*, Kerman, Iran. (In Farsi)
- Gavin H. and Agnew C. A. (2004). Modelling actual reference and equilibrium evaporation from a temperate wet grassland. *Hydrological Processes*, 18, 229-246.
- Ghahreman N. and Gharekhani A. (2011). Evaluation of random time series models in estimating pan evaporation (case study: Shiraz station). *Journal of Water Research in Agriculture*, 25(1), 75-81.
- Gundekar H. G., Khodke U. M. and Sarkar S. (2008). Evaluation of pan coefficient for reference crop evapotranspiration for semi-arid region. *Irrigation Science*, 26, 169-175.
- Hassan M. (2013). Evaporation estimation for Lake Nasser based on remote sensing technology. *Ain Shams Engineering Journal*, 4, 593-604.
- Khalili Naft Chali A., Khashei Siuki A. and Shahidi A. (2017). Compare KNN and M5 decision tree models in anticipation of evaporation and comparison with empirical equations (Case Study of Birjand). *Iranian Journal of Irrigation & Drainage*, 11(3), 356-366.
- Kuss M. (2006). *Gaussian process models for robust regression, classification, and reinforcement learning*. Ph. D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany.
- Majidi M., Alizadeh A., FaridHosseini A. and Vazifedoust M, (2014). Lake and reservoir evaporation: energy balance estimations, evaluation of combination and radiation-temperature methods. *Iranian Journal of Irrigation and Drainage*, 3(8), 602-615. (In Farsi)
- McGuinness J.L., Bordn, E.F. (1972). *A comparison of lysimeterderived potential evapotranspiration with computed values*. Technical Bulletin 1452, US Department of Agriculture Agricultural Research Service, Washington, DC.
- Mouneskhah V., Majnooni-Heris A. and Fakheri-Fard A. (2018). Evaluation and calibration of empirical relationships for estimating evaporation from free water levels in Urmia Lake Basin. *Iranian Journal of Irrigation and Drainage*, 5(12), 1281-1291. (In Farsi)
- Qasem S., Samadianfard S., Kheshtgar S., Jarhan S., Kisi O., Shamshirband SH. and Wing-Chau K. (2019). Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Engineering Applications of Computational Fluid Mechanics*, 13(1), 177-187.
- Rosenberry D.O., Winter T.C., Buso D.C., and Likens G.E. (2007). Comparison of 15 evaporation

- methods applied to a small mountain lake in the northeastern USA. *Journal of Hydrology*, 340, 149–166.
- Samadianfard S., Hashemi S., and Izadyar M. (2018). Estimation of daily pan evaporation by using machine learning methods. *Iranian Journal of Irrigation and Drainage*, 4(12), 1004-1015. (In Farsi)
- Shabani S., Samadianfard S., Sattari M.T., Mosavi A., Shamshirband Sh., Kmet T. and Annamaria R. (2020). Modeling pan evaporation using gaussian process regression k-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, 11(66), 1-17.
- Shadmani M. and Marofi S. (2011). Comparison of some methods for estimation of daily pan evaporation: case study in Kerman Region. *Journal of Science and Technology of Agriculture and Natural Resources*, 15(55), 69-83. (In Farsi)
- Sharifazari S. and Araghinejad S. (2013). Develop a non-parametric model to simulate monthly hydrological data. *Water and Irrigation Management*, 3(1), 83-95. (In Farsi)
- Singh D., Ganju A. and Singh A. (2005). Weather prediction using nearest-neighbor model. *Current science*, 88, 8-25.
- Singh A. K., Tripathy R., and Chopra U. K. (2008). Evaluation of CERESWheat and CropSystmodels for water-nitrogen interactions in wheat crop. *Agricultural Water Management*, 95, 776-786.
- Sun Z., Wei B., Su W., Shen W., Wang C., You D and Liu Z. (2011). Evapotranspiration estimation based on the SEBAL model in the Nansi Lake Wetland of China. *Mathematical and Computer Modelling*, 54(3), 1086-1092.
- Tabari H., Marufi S., and Sabziparvar A.A. (2010). Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigation Science*, 28(3), 399-406.
- Terzi O. (2011). Modeling of daily pan evaporation of Lake Egirdir using data-driven techniques. *International symposium on innovations in intelligent systems and Applications*, Istanbul, Turkey, pp. 320-324.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York. 314 pp.
- Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley, New York. 736 pp.
- Wu X., Kumar V., Quinlan J.R., Ghosh J., Yang Q., Motoda H., McLachlan G.J., Ng A., Liu B. and Philip S.Y. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1–37.