



Research Paper

Evaluating Capabilities of Gradient Boosted Tree and Optimized Random Forest Models in Estimating Daily Dew Point Temperature

Mohsen Osouli Shojaei¹, Fatemeh Mikaeili¹ and Saeed Samadianfard^{2*}

¹M.Sc. Student, Department of Water Engineering, Faculty of Agriculture, Tabriz University, Tabriz, Iran

²Assoc. Professor, Department of Water Engineering, Faculty of Agriculture, Tabriz University, Tabriz, Iran

Article information

Received: November 02, 2021

Revised: January 12, 2022

Accepted: January 12, 2022

Keywords:

Ardabil

Intelligence Models

Meteorological Variables

Statistical Evaluation

*Corresponding author:

s.samadian@tabrizu.ac.ir



Abstract

Dew point temperature is very important in various fields including meteorology for weather forecasts. Therefore, it is necessary to provide suitable models to accurately predict the value of this meteorological variable for the practical use of agricultural engineers and nearby stations where it is not possible to measure this temperature. In the present study, we investigated the ability of four data-driven models, including gradient reinforcement tree, MSP tree model, random forest, and random forest optimized with genetic algorithm, in estimating daily dew point temperature. For this purpose, the daily meteorological data of two stations in Ardabil and Parsabad were used in the period of 2014 to 2019. The used meteorological parameters include minimum, maximum, and average temperature, relative humidity, sunshine hour, and wind speed, which were considered input variables for each of the mentioned models in 10 different combinations. The comparison of the results obtained for both stations showed that the MSP-8 model with a root mean square error of 0.54°C and a Wilmot coefficient equal to 0.998 in the Ardabil station and the MSP-6 model with a root mean square error of 0.29°C and Wilmot coefficient equal to 1.00 was introduced as the best models in Parsabad station.

© Authors, Published by **Environment and Water Engineering** journal. This is an open-access article distributed under the CC BY (license <http://creativecommons.org/licenses/by/4.0/>).



Introduction

Dew point temperature (DPT) is the temperature at which the moisture or vapor in the air is converted to liquid water at a constant pressure due to the high concentration of water molecules. Accurate estimation of this temperature plays an important role in estimating frost, snow, rain, humidity, and other meteorological parameters. Long-term changes in dew point temperature can be very effective in regulating the energy adjacent to the earth's surface and the water

balance and greenhouse gases. DPT is not measured and recorded relative to common meteorological parameters such as rainfall and relative humidity at many meteorological stations, but is mainly estimated using regression methods, which in most cases estimated data are not reliable. Today, intelligent systems based on data mining have been considered in modeling hydrological and meteorological processes. According to the conducted research, the importance of using data-driven methods in



estimating the dew point temperature is determined.

Moreover, the review of the sources showed that the evaluation of the abilities of the mentioned models in estimating the dew point temperature has not been done so far. Therefore, in this research, using different meteorological data in different combinations as input to M5P tree models, Gradient boosted tree (GBT), random forest (RF) and random forest optimized with genetic algorithm (RF-GA), dew point temperature in two synoptic stations of Ardabil province will be estimated. Also, using a genetic algorithm, the values obtained from RF models will be optimized. Finally, by comparing the results of different scenarios in the two studied stations, the most accurate model will be selected and introduced in each station. Finally, the mathematical equations provided to accurately estimate this temperature in nearby stations where it is not possible to measure and record the dew point temperature can be used by researchers and agricultural engineers.

Material and Methods

In the present study, daily meteorological data of Ardabil and Parsabad stations including mean temperature, minimum temperature, maximum temperature, relative humidity, sunshine hours, wind speed, and dew point temperature were used for a period of 16 years (2005-2011). Based on this, 10 combined scenarios of meteorological variables were defined as the input of the studied models. These combinations were defined based on the correlation coefficient between the dew point temperature parameter of each station and other meteorological parameters of that station. The methods used in the present study include GBT, M5P, RF, and RF-GA. In this research, 70% of the data is considered for training and 30% for testing. The results obtained from these models were compared using statistical indices such as coefficient of determination (R^2), root mean square error (RMSE), and Willmott's index (WI), and the best models were selected.

Results

The results showed that at the Ardabil station, M5P-8 and M5P-10 models with $R^2 = 0.994$, $WI = 0.998$ and the $RMSE = 0.54$ °C had the best performance compared to the different patterns defined for each of the mentioned models. the

second place, the RF-GA-6 model showed remarkable performance compared to other defined patterns and methods. Pattern No. 6 with the input parameters of temperature and relative humidity in all methods except M5P, has provided the most accurate results. In Ardabil station, model RF-1 with $R^2 = 0.706$, $WI = 0.91$ and the $RMSE = 3.79$ °C showed the weakest performance among the other methods. At Parsabad station, the M5P model in scenarios 6, 8, 9, and 10 with the same results with $R^2 = 0.999$, $WI = 1.00$ and $RMSE = 0.29$ °C had the best prediction and the performance of the GA-RF-6 model was ranked second. The GBT-1 model with a $R^2 = 0.908$, $WI = 0.78$ and $RMSE = 4.38$ °C presented the weakest results. In general, the M5P tree model in both stations was introduced as the superior model in estimating dew point temperature values, and the GBT model with poor results was considered an inefficient method in predicting daily dew point temperature values. In Parsabad station, the methods used due to the high correlation of temperature and relative humidity values with dew point temperature, the accuracy of estimations was higher and had better performance (high coefficient of determination and low error) compared to Ardabil station, and it should be noted that the temperature parameter was a necessary but insufficient condition in estimating the dew point temperature.

Although the correlation value of relative humidity with dew point temperature is low, probably the relative humidity parameter in linear combination with temperature parameters has a high correlation with dew point temperature and according to the comparison of the mentioned results, with the addition of relative humidity to the input data, the accuracy of the models increases significantly. In Fig. 1 the root mean square error values of the proposed models can be seen for each defined pattern in both stations.

As can be seen from Fig. 1, M5P, RF-GA and RF models have low error and GBT model has a high error, and pattern number 6 has the lowest error among the different defined patterns with parameters of temperature and relative humidity.

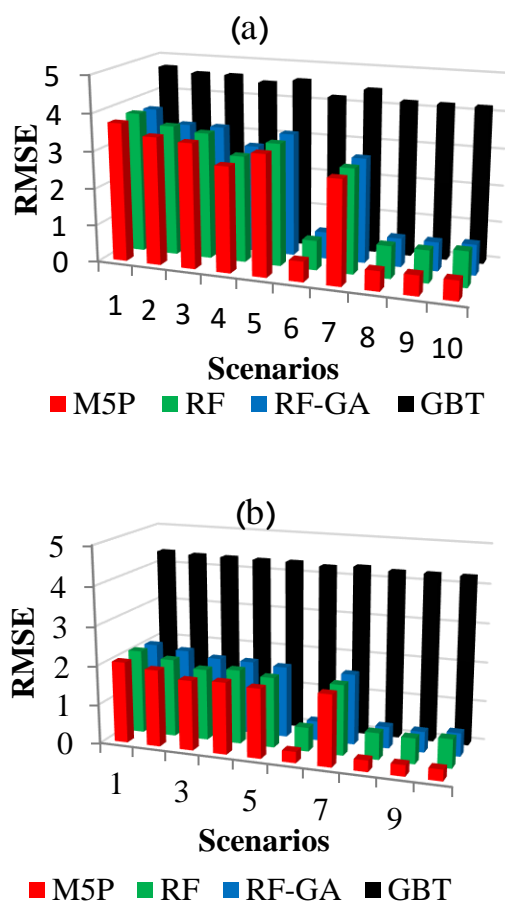


Fig. 1 RMSE values of the proposed models per scenario for: a) Ardabil and b) Parsabad stations

Conclusion

One of the applications of dew point temperature is in the science of ventilation as well as meteorology, which is very important in weather forecasts. Therefore, in the current research, it was tried to estimate the dew point temperature using M5P, RF, RF-GA, and GBT methods for two weather stations in northwest of Iran, and the most accurate method was suggested. Finally, by

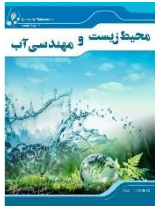
performing statistical analysis, the overall conclusion showed that: 1) In Ardabil station, model M5P in the eighth scenario includes input data of relative humidity, wind speed, minimum, maximum, and average temperature and in ParsAbad station, the model in the sixth scenario includes input data of relative humidity, minimum, maximum and average temperature, the most desirable results offered. 2) Comparing the results between the models, the M5P, GA-RF, RF, and GBT models were ranked from more to less accuracy, respectively. 3) Parameters of temperature and relative humidity were introduced as the dominant variables in predicting dew point temperature. 4) By comparing the results of RF with GA-RF models, it was concluded that the genetic algorithm improved the performance of random forest and had a positive effect on all models. Finally, as a general result, it can be stated that the M5P model had a high ability to estimate the daily values of the dew point temperature and therefore, it is recommended in the field of meteorological forecasts and agricultural activities.

Data Availability

The data can be sent on request by the corresponding author via s.samadian@tabrizu.ac.ir email.

Conflicts of interest

The authors of this article declared no conflict of interest regarding the authorship or publication of this article.



ISSN: 2476-3683

محیط‌زیست و مهندسی آب

Homepage: www.jewe.ir

مقاله پژوهشی

ارزیابی توانایی مدل‌های گرادیان تقویتی و جنگل تصادفی بهینه شده در تخمین دمای نقطه شبنم روزانه

محسن اصولی شجاعی^۱، فاطمه میکائیلی^۱ و سعید صمدیان فرد^{۲*}

^۱دانشجوی کارشناسی ارشد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

^۲دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

اطلاعات مقاله

چکیده

تاریخ دریافت: [۱۴۰۰/۰۸/۱۱]

تاریخ بازنگری: [۱۴۰۰/۱۰/۲۲]

تاریخ پذیرش: [۱۴۰۰/۱۰/۲۲]

واژه‌های کلیدی:

اردبیل
ارزیابی آماری
متغیرهای هواشناسی
مدل‌های هوشمند

*نویسنده مسئول:

s.samadian@tabrizu.ac.ir



دمای نقطه شبنم در زمینه‌های مختلف از جمله علوم هواشناسی جهت پیش‌بینی‌های مربوط به آب و هوا دارای اهمیت فراوانی می‌باشد. لذا ارائه مدل‌های مناسب جهت پیش‌بینی دقیق مقدار این متغیر هواشناسی برای استفاده عملی مهندسی کشاورزی و ایستگاه‌های مجاوری که در آن‌ها امکان اندازه‌گیری این دما وجود ندارد، ضروری می‌باشد. در پژوهش حاضر توانایی چهار مدل داده‌محور شامل درخت گرادیان تقویتی، مدل درختی MSP، جنگل تصادفی و جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک در تخمین دمای نقطه شبنم روزانه مورد ارزیابی قرار گرفت. برای این منظور از داده‌های هواشناسی روزانه دو ایستگاه اردبیل و پارس‌آباد در بازه زمانی ۱۳۸۴ تا ۱۳۹۹ استفاده شد. پارامترهای هواشناسی مورد استفاده شامل حداقل، حداکثر و میانگین دما، رطوبت نسبی، ساعت آفتابی و سرعت باد بوده که در ۱۰ ترکیب متفاوت به‌عنوان متغیرهای ورودی برای هر یک از مدل‌های مذکور در نظر گرفته شدند. مقایسه نتایج به‌دست آمده برای هر دو ایستگاه نشان داد که مدل MSP-8 با دارا بودن جذر میانگین مربعات خطای 0.54°C و ضریب وایلموت برابر با ۰/۹۹۸ در ایستگاه اردبیل و مدل MSP-6 با جذر میانگین مربعات خطای 0.29°C و ضریب وایلموت برابر با ۱/۰۰ در ایستگاه پارس‌آباد به‌عنوان برترین مدل‌ها معرفی شدند.

۱- مقدمه

دمای نقطه شبنم می‌تواند در تنظیم انرژی مجاور سطح زمین و تعادل آبی و گازهای گلخانه‌ای بسیار مؤثر باشد (Deka et al. 2017). دسترسی دقیق و مطمئن به این دما در بسیاری از تحقیقات علمی مربوط به مسائل کشاورزی و حوزه‌های مختلف علوم آب دارای اهمیت می‌باشد (Amirmojahedi et al. 2016). اندازه‌گیری و ثبت دمای

دمای نقطه شبنم دمایی است که در آن رطوبت یا بخار موجود در هوا به دلیل غلظت زیاد مولکول‌های آب، در فشار ثابت به آب مایع تبدیل شود (Shank et al. 2008). برآورد دقیق این دما نقش مهمی در تخمین سرمازدگی، یخ‌زدگی، برف، باران، میزان رطوبت موجود در هوا و سایر پارامترهای هواشناسی دارد (Shiri et al. 2014). تغییرات بلندمدت



تمام مکان‌های مورد مطالعه ارائه داده‌اند. Qasem et al. (2019) در تحقیق خود با هدف تخمین دمای نقطه شبنم در شهر تبریز به کمک سه مدل GEP، مدل درختی M5 و رگرسیون بردار پشتیبان به این نتیجه رسیدند که مدل M5 در مقایسه با سایر مدل‌های مورد مطالعه نتایج بهتری ارائه کرده است. Naganna et al. (2019) پژوهشی را با هدف بررسی کارایی شبکه عصبی پرسپترون چند لایه (MLP)^۵، جستجوی گرانشی و کرم شبتاب، در برابر MLP استاندارد بهینه شده به وسیله الگوریتم پس انتشار مارکوارت-لونبرگ، ماشین یادگیری ELM و مدل‌های ماشین بردار پشتیبان (SVM) جهت تخمین دمای نقطه شبنم به انجام رساندند. آن‌ها گزارش نمودند که مدل‌های ترکیبی MLP پیشنهادی دقت برآورد بیشتری دارند. Alizamir et al. (2020) در پژوهش خود به منظور برآورد دمای نقطه شبنم، مدل‌های دستگاه یادگیری شدید کرنل، درخت رگرسیون تقویت شده، شبکه عصبی تابع پایه شعاعی، شبکه عصبی پرسپترون چند لایه و رگرسیون کم‌انکلی تطبیقی چند متغیره را مورد بررسی قرار دادند. مقایسه نتایج عملکرد مدل‌های پیشنهادی نشان داد مدل دستگاه یادگیری شدید کرنل در هر دو ایستگاه، برآورد بهتری داشته است. Dong et al. (2020) در مطالعه‌ای به مقایسه توانایی مدل الگوریتم خفاش بهینه شده توسط ماشین یادگیری شدید در مقابل مدل غیرخطی نرخ کاهش تولیدی آرپس بر پایه کرنل (هسته) و چند مدل دیگر پرداختند. نتایج مطالعه آن‌ها نشان داد که Bat-ELM در بین مدل‌های دیگر دمای نقطه شبنم روزانه را دقیق‌تر برآورد کرده است.

با توجه به پژوهش‌های انجام گرفته، اهمیت کاربرد روش‌های داده محور در برآورد دمای نقطه شبنم مشخص می‌شود. همچنین بررسی منابع نشان داد که تاکنون ارزیابی توانایی مدل‌های مذکور در تخمین دمای نقطه شبنم به انجام نرسیده است. لذا در این پژوهش با استفاده از داده‌های مختلف هواشناسی در ترکیب‌های مختلف به‌عنوان ورودی مدل‌های درختی M5P، درخت گرادیان تقویتی (GBT)^۶، جنگل تصادفی (RF) و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (RF-GA)، دمای نقطه شبنم در دو ایستگاه هواشناسی استان اردبیل برآورد خواهد شد. همچنین با به‌کارگیری الگوریتم ژنتیک مقادیر به‌دست‌آمده از مدل‌های

نقطه شبنم نسبت به پارامترهای متداول هواشناسی مانند بارندگی و رطوبت نسبی در بسیاری از ایستگاه‌های هواشناسی انجام نمی‌گیرد، بلکه عمدتاً با استفاده از روش‌های رگرسیونی تخمین زده می‌شود که در بیشتر مواقع داده‌های تخمینی قابل اعتماد نیستند (Dong et al. 2020). در سال‌های اخیر مدل‌های داده محور از جمله مدل‌های درختی جهت تخمین متغیرها و مسائل هیدرولوژیکی بیشتر مورد توجه پژوهشگران واقع شده است. در این زمینه مطالعات متعددی انجام گرفته است از جمله Shiri et al. (2014) با مقایسه عملکرد دو مدل مختلف شبکه عصبی مصنوعی (ANN)^۱ و برنامه‌ریزی بیان ژن^۲ (GEP) برای تخمین دمای نقطه شبنم چنین اظهار داشتند که مدل GEP نسبت به ANN این دما را دقیق‌تر پیش‌بینی کرده است. Mohammadi et al. (2016) در کار تحقیقی خود سیستم استنتاج عصبی فازی-تطبیقی (ANFIS) را به‌عنوان روشی جهت شناسایی مهم‌ترین پارامترها برای پیش‌بینی دمای نقطه شبنم روزانه بررسی نمودند. نتایج حاکی از آن بود که علی‌رغم تفاوت شرایط آب و هوایی دو ایستگاه، فشار بخار آب تأثیرگذارترین پارامتر است. همچنین مدل ANFIS دقت بالاتری در مقایسه با پیش‌بینی سایر روش‌های محاسبه نرم داشته است. Amirmojahedi et al. (2016) یک روش ترکیبی ماشین یادگیری شدید (ELM)^۳ و الگوریتم تبدیل موجک (WT) برای پیش‌بینی دمای نقطه شبنم روزانه در شهر بندرعباس را پیشنهاد کردند. آن‌ها اظهار داشتند که روش پیشنهادی ELM-WT در مقابل ELM، ماشین‌های بردار پشتیبان و فن‌های شبکه عصبی مصنوعی دقیق‌تر عمل کرده است.

Shiri (2018) در شش ایستگاه هواشناسی مختلف در استان آذربایجان شرقی جهت توسعه و اعتبارسنجی مدل‌های پیش‌بینی و تخمین مقادیر دمای نقطه شبنم از روش‌های GEP، مدل رگرسیون کم‌انکلی تطبیقی چند متغیره و مدل جنگل تصادفی (RF)^۴ استفاده کردند. نتایج به‌دست آمده نشان داد که مدل GEP از مدل‌های مذکور دقت بالاتری داشته است. همچنین در میان مدل‌های تخمینی نیز، مدل‌های رگرسیون کم‌انکلی تطبیقی دقیق‌ترین نتایج را در

¹Artificial Neural Network

²Genetic Programming

³Extreme Learning Machine

⁴Random Forest

⁵Multilayer Perceptron

⁶Gradient Boosted Tree



۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

در پژوهش حاضر از داده‌های هواشناسی دو ایستگاه سینوپتیک اردبیل و پارس‌آباد واقع در استان اردبیل استفاده شده است. اطلاعات جغرافیایی ایستگاه‌های مورد مطالعه در جدول (۱) قابل ملاحظه است.

RF بهینه‌سازی خواهد شد. در نهایت با مقایسه نتایج سناریوهای مختلف در دو ایستگاه مورد مطالعه دقیق‌ترین مدل در هر ایستگاه انتخاب و معرفی خواهد شد. همچنین معادلات ریاضی ارائه شده جهت تخمین دقیق این دما در ایستگاه‌های مجاور که امکان اندازه‌گیری و ثبت دمای نقطه شب‌نم در آن‌ها وجود ندارد می‌تواند مورد استفاده پژوهشگران و مهندسين کشاورزی قرار گیرد.

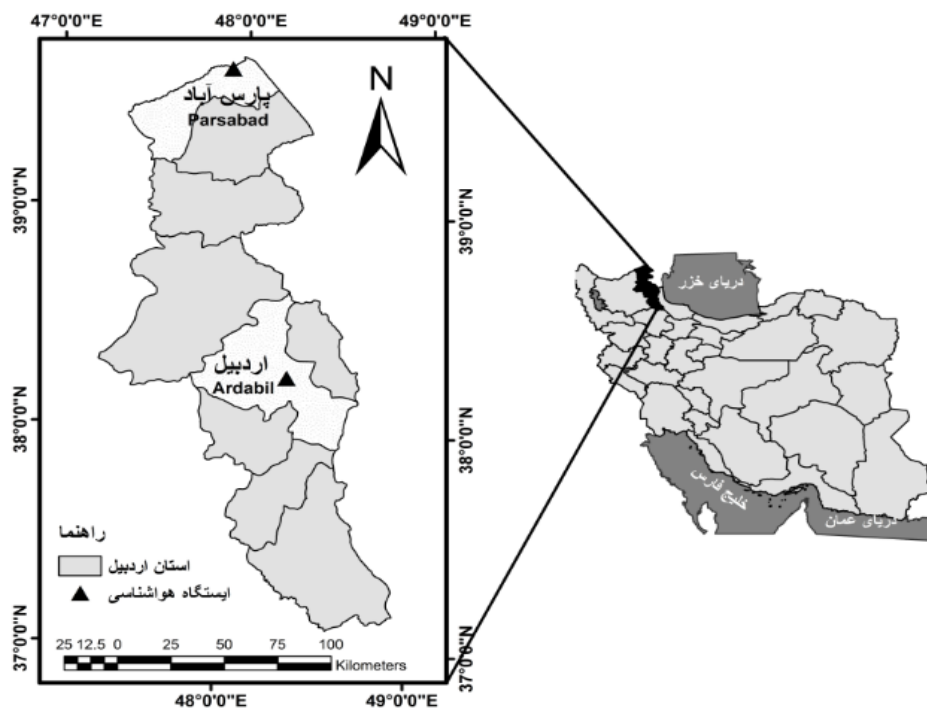
جدول ۱- مشخصات جغرافیایی ایستگاه‌های مورد مطالعه

Table 1 Geographical profile table of the studied station

Station	Latitude	Longitude	Elevation from sea level (m)	Climate
Ardabil	38° 15'	48° 17'	1335	Semi-arid with mild summer and very cold winter
Parsabad	39° 39'	47° 55'	72.6	Semi-arid with hot summer and cold winter

باعث شده این استان از لحاظ میزان نزولات جوی وضعیت بهتری نسبت به سایر مناطق کشور داشته باشد. موقعیت جغرافیایی منطقه مورد مطالعه در شکل (۱) نشان داده شده است.

استان اردبیل با وسعتی حدود 17800 km^2 تقریباً ۱/۱٪ از مساحت کل کشور ایران را به خود اختصاص داده است. ویژگی‌های خاص جغرافیایی و توپوگرافی استان مانند رشته‌کوه‌هایی با ارتفاع بیش از ۴۰۰۰ m و دشت‌های وسیع



شکل ۱- موقعیت جغرافیایی استان اردبیل و ایستگاه‌های مورد مطالعه

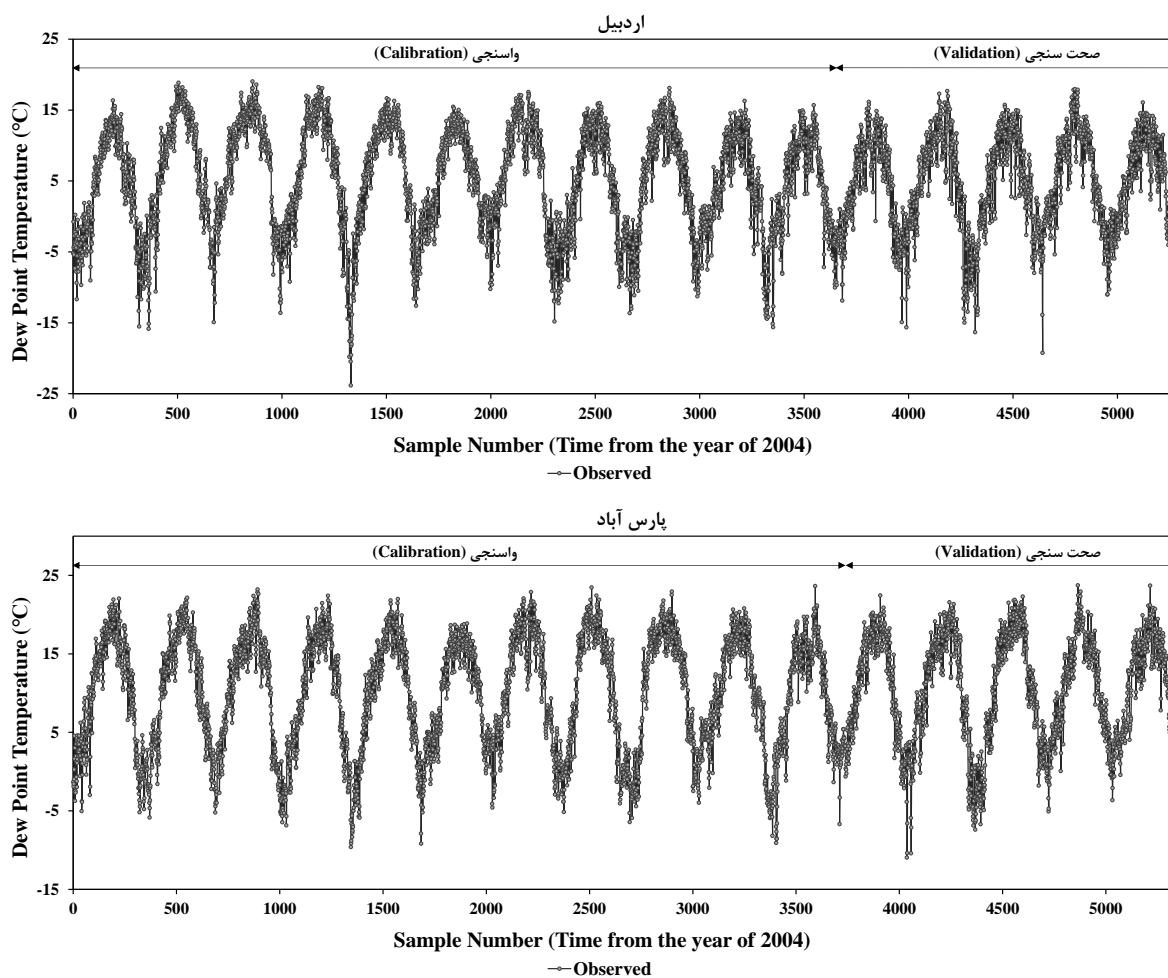
Fig. 1 Geographical location of Ardabil province and studied stations

(DPT) برای یک دوره ۱۶ ساله (۱۳۸۴ تا ۱۳۹۹) می‌باشند. بدین منظور، مشخصات آماری پارامترهای مذکور در هر دو مرحله اعتبارسنجی و صحت‌سنجی در ایستگاه‌های مورد مطالعه در جدول (۲) ارائه شده است.

داده‌های هواشناسی مورد استفاده در پژوهش حاضر از سازمان هواشناسی استان اردبیل اخذ شد. این داده‌ها در مقیاس روزانه شامل دمای میانگین (T)، دمای کمینه (T_{min})، دمای بیشینه (T_{max})، رطوبت نسبی (RH)، ساعات آفتابی (S)، سرعت باد (W) و همچنین دمای نقطه شب‌نم

جدول ۲- مشخصات آماری داده‌ها در هر دو مرحله آموزش و آزمون
Table 2 Statistical characteristics of the data in both training and testing stages

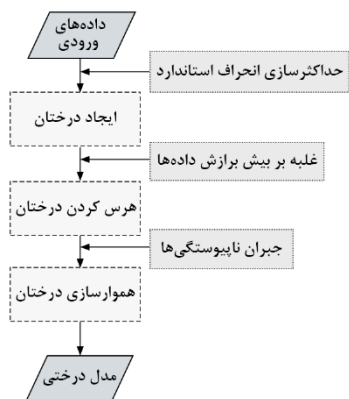
	Training Stage							Testing Stage						
	Variable	unit	X _{mean}	X _{min}	X _{max}	S _x	C _{sx}	Variable	unit	X _{mean}	X _{min}	X _{max}	S _x	C _{sx}
Ardabil	T	°C	10.44	-22.4	27.5	8.23	-0.59	T	°C	11.38	-15.3	29.9	7.85	-0.42
	T _{min}	°C	4.39	-3.14	19.4	7.86	-0.78	T _{min}	°C	5.51	-23.0	20	7.12	-0.56
	T _{max}	°C	17.05	-11.8	39.2	9.29	-0.45	T _{max}	°C	18.19	-7.00	40.4	9.12	-0.32
	W	m/s	3.53	0.00	16.63	2.42	1.13	W	m/s	3.61	0.00	13.5	2.51	1.11
	S	h	7.35	0.00	13.7	3.67	-0.50	S	h	7.60	0.00	13.5	2.51	-0.59
	RH	%	73.33	20.00	100.0	14.94	-0.55	RH	%	69	17.88	100	16.61	-0.18
Parsabad	T	°C	15.64	-6.9	32.2	9.08	-0.13	T	°C	16.16	-8.2	31.4	8.9	-0.13
	T _{min}	°C	10.51	-13	27.4	8.33	-0.18	T _{min}	°C	10.85	-13.2	27.3	8.09	-0.18
	T _{max}	°C	21.72	-2.8	40.6	9.99	-0.16	T _{max}	°C	22.66	-1.3	44.0	10.03	-0.15
	W	m/s	1.96	0.00	10.13	1.45	1.08	W	m/s	2.91	0.00	13.38	1.44	1.50
	S	h	6.58	0.00	13.7	4.11	-0.32	S	h	6.46	0.00	16.0	3.98	-0.34
	RH	%	72.02	32.25	97.5	11.95	-0.29	RH	%	73.09	36.13	99.75	12.63	-0.28



شکل ۲- نمودار تغییرات دمای نقطه شبنم روزانه طی دوره آماری ۱۳۹۹-۱۳۸۴ در ایستگاه‌های اردبیل و پارس آباد
Fig. 2 Fluctuations of daily dew point temperature in time period of 2004-2019 in Ardabil and Parsabad stations



اول با داده‌هایی که به اشتباه توسط درخت پیشین پیش‌بینی شده‌اند، آموزش داده می‌شوند. این امر موجب می‌شود مدل بیشتر روی موارد پیچیده و کمتر بر مسائلی که پیش‌بینی در آن‌ها آسان است، متمرکز شود. بنابراین این روش نسبت به روش رگرسیون خطی و روش دسته‌بندی^۳ نتایج بهتری دارد (Siasar et al. 2020).



شکل ۳- فلوجارت مدل درختی M5P (Zhan et al. 2011)
Fig. 3 Flowchart of the M5P tree flowchart

۲-۴- مدل جنگل تصادفی RF

الگوریتم جنگل تصادفی به‌عنوان یک روش غیر پارامتریک یادگیری ماشین متعلق به خانواده روش‌های دسته‌جمعی است. این روش، برخلاف مدل‌های کلاسیک مانند رگرسیون که نه تنها بر پایه یک مدل تکیه دارند با استفاده از تعداد زیادی درخت تصمیم، از اطلاعات بیش‌تری در داده‌ها استفاده می‌کند. استفاده از چندین درخت تصمیم‌گیری، بی‌ثباتی و حساسیت مدل را کاهش داده و منجر به افزایش قدرت پیش‌بینی مدل خواهد شد (Breiman 2001).

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چند نمونه Bootstrap از داده‌ها شرکت دارند، در ساخت هر درخت به‌طور تصادفی تعدادی از متغیرهای ورودی برای جلوگیری و کاهش تطابق میان درخت‌های تصمیم شرکت می‌کنند و هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی کننده‌ها نسبتاً کم باشد یک روش پیش‌بینی کارآمد به حساب می‌آید (Breiman 2001).

مدل‌سازی با استفاده از روش‌های داده‌محور شامل مدل درختی M5P، درخت گرادیان تقویتی (GBT)، جنگل تصادفی (RF) و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (RF-GA) انجام شده است. در این پژوهش ۷۰٪ داده‌ها برای آموزش و ۳۰٪ برای آزمون در نظر گرفته شده است. در شکل (۲) نمودارهای تغییرات زمانی دمای نقطه شب‌نم روزانه در طی دوره آماری ۱۳۹۹-۱۳۸۴ در دو ایستگاه مورد مطالعه نشان داده شده است.

۲-۲- مدل درختی M5P

روش‌های مبتنی بر درخت تصمیم به‌عنوان یکی از شناخته شده‌ترین فن‌های داده‌کاوی، ویژگی هدف را به‌عنوان خروجی به‌صورت یک مدل با سازه درختی با استفاده از داده‌های ورودی پیش‌بینی و یا طبقه‌بندی می‌نمایند (Qasem et al. 2019). در مرحله اول از الگوریتم استنباطی یا معیار انشعاب برای تولید درخت استفاده می‌شود. معیار انشعاب برای مدل M5P براساس عملکرد انحراف استاندارد مقادیر هر کلاس و یا طبقه است که در هر گره به‌دست آمده است. معیار انشعاب بیانگر میزان خطا در آن گره می‌باشد و مدل حداقل خطای مورد انتظار را به‌عنوان نتیجه آزمایش هر صفت در آن گره محاسبه می‌کند. فرمول محاسبه نسبت انحراف استاندارد^۱ (SDR) به‌صورت رابطه‌های (۱) و (۲) می‌باشد.

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (1)$$

$$sd(T) = \sqrt{\sum_1^n \frac{(T_i - \bar{T})^2}{n-1}} \quad (2)$$

که، T مجموعه‌ای از نمونه‌هاست که به هر گره وارد می‌شود، T_i زیرمجموعه‌ای از نمونه‌هاست که i مین نتیجه آزمون بالقوه را دارند، sd انحراف معیار و n تعداد نمونه‌ها می‌باشند (Qasem et al. 2019).

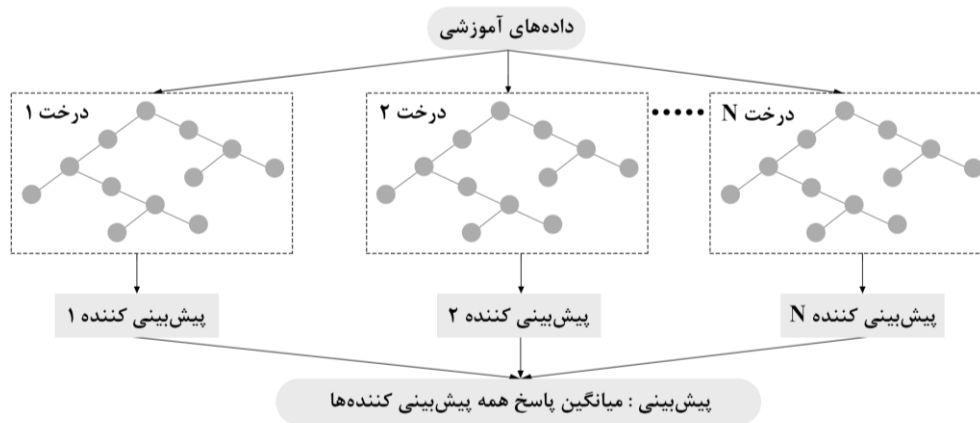
۲-۳- مدل درخت گرادیان تقویتی GBT

روش درخت گرادیان تقویتی (GBT)^۲ یکی از جدیدترین الگوریتم‌های یادگیری است و برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با دقت بالایی انجام می‌دهد. در این روش، درخت‌ها یکی پس از دیگری آموزش داده می‌شوند. هر درخت زیرمجموعه، در درجه

¹Standard Deviation Ratio

²Gradient Boosted Tree

³Bagging



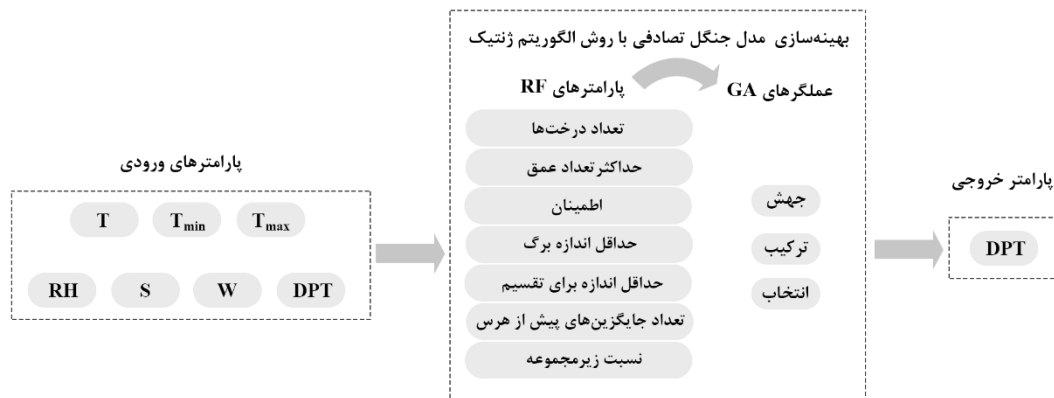
شکل ۴- نمای کلی مدل روش جنگل تصادفی

Fig. 4 Overview of the RF model

برای یک مجموعه طبقه‌بندی شده خاص $h_1(x), h_2(x), \dots, h_k(x)$ توابع حاشیه را می‌توان به صورت رابطه (۳) نوشت (Samadianfard et al. 2020).

$$mg(X, Y) = a\vartheta_k I(h_k(X) = Y) - \max_{j \neq Y} a\vartheta_k I(h_k(X) = j) \quad (3)$$

که، I تابع شاخص است و پیش‌بینی نهایی با میانگین حاصل از همه درختان مشخص می‌شود.



شکل ۵- نمای کلی مدل ترکیبی RF-GA

Fig. 5 Overview of the RF-GA model

ارتقای بهره‌وری دقت مدل جنگل تصادفی بهینه‌سازی پارامترهای کلیدی در ساختار مدل RF یک کار ضروری است (شکل ۵) که این روش شامل دو مرحله اصلی است: تنظیم پارامتر و بهینه‌سازی. در مرحله بعد، بهینه‌سازی با استفاده از الگوریتم ژنتیک برای یافتن ترکیبی بهینه از درخت‌های تصمیم با هدف به حداکثر رساندن امتیاز بهره با بررسی بازده و تلفات واقعی و بالقوه انجام می‌شود (Samadianfard et al. 2021).

۲-۵- مدل جنگل تصادفی بهینه شده با الگوریتم ژنتیک RF-G

الگوریتم ژنتیک (GA)^۱ با الهام از ژنتیک و انتخاب طبیعی، فنی برای بهینه‌سازی و جستجو مستقیم است که با در نظر گرفتن مجموعه‌ای از نقاط فضای جواب در هر تکرار محاسباتی، به نحو مؤثری نواحی مختلف فضای جواب را جستجو می‌کند. در الگوریتم ژنتیک دو نوع عملگر شامل عملگرهای تکاملی مانند انتخاب و عملگرهای ژنتیک مانند ترکیب و جهش وجود دارد. (Dong et al. 2020). برای

^۱Genetic Algorithm

محاسبات صحت‌سنجی بر روی ۳۰٪ از داده‌های باقی‌مانده، انجام شد.

جدول ۴- الگوهای مختلف ترکیب پارامترهای ورودی در برآورد دمای نقطه شب‌نم

Table 4 Different combinations of input parameters in the estimation of dew point temperature

Scenario Number	Input Parameters
1	T
2	T _{min} , T _{max}
3	T, T _{min} , T _{max}
4	T, T _{min} , T _{max} , W
5	T, T _{min} , T _{max} , S
6	T, T _{min} , T _{max} , RH
7	T, T _{min} , T _{max} , W, S
8	T, T _{min} , T _{max} , W, RH
9	T, T _{min} , T _{max} , S, RH
10	T, T _{min} , T _{max} , W, S, RH

با تعریف ۱۰ الگو از پارامترهای هواشناسی، محاسبات پیش‌بینی دمای نقطه شب‌نم با روش‌های RF، M5P، GBT و RF-GA انجام گرفت. دقت روش‌های مذکور در مرحله صحت‌سنجی بر اساس معیارهای آماری RMSE، WI و R² (روابط ۳، ۴ و ۵) محاسبه شده و نتایج در جدول (۵) ارائه گردید. با توجه به نتایج ارائه شده در جدول (۵)، در ایستگاه اردبیل مدل M5P-8 و M5P-10 با ضریب تعیین ۰/۹۹۴، ضریب توافق ویلموت ۰/۹۹۸ و جذر میانگین مربعات خطای برابر با ۰/۵۴ بهترین عملکرد را نسبت به الگوهای مختلف تعریف شده برای هر یک از مدل‌های مذکور داشت. در مرتبه دوم نیز مدل RF-GA-6 عملکرد قابل‌توجهی را در مقایسه با سایر الگوها و روش‌های تعریف شده از خود نشان داد. الگوی شماره ۶ با پارامترهای ورودی دما و رطوبت نسبی در تمامی روش‌ها به جز روش M5P (با اختلاف اندک از الگوهای ۸، ۹ و ۱۰)، بهترین نتایج را ارائه کرده است و گزینه مطلوبی از نظر تعداد داده می‌باشد. در ایستگاه اردبیل مدل RF-1 با ضریب تعیین ۰/۷۰۶، ضریب توافق ویلموت ۰/۹۱ و جذر میانگین مربعات خطای برابر با ۳/۷۹ ضعیف‌ترین عملکرد را در بین سایر روش‌های مورد مطالعه از خود نشان داد و همچنین لازم به ذکر است که الگوی شماره ۱ تنها با پارامتر ورودی دمای متوسط ضعیف‌ترین عملکرد را در همه روش‌ها ارائه داده است. در ایستگاه پارس‌آباد مدل M5P در الگوهای ۶، ۸، ۹ و ۱۰ با ارائه نتایج مشابه با ضریب تعیین ۰/۹۹۹، ضریب توافق ویلموت ۱/۰۰ و جذر میانگین

۲-۶- معیارهای ارزیابی عملکرد مدل‌ها

جهت ارزیابی دقت نتایج و مقایسه قدرت پیش‌بینی مدل‌های پیشنهادی از شاخص‌های جذر میانگین مربعات خطا (RMSE)^۱ (رابطه ۳)، شاخص توافق ویلموت (WI)^۲ (رابطه ۴) و ضریب تعیین (R²)^۳ (رابطه ۵) استفاده شد.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

$$WI = 1 - \left(\frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |x_i - \bar{x}|)^2} \right) \quad (4)$$

$$R^2 = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \right)^2 \quad (5)$$

که، y_i مقدار برآورد شده دمای نقطه شب‌نم به کمک مدل در روز i، x_i مقدار دمای نقطه شب‌نم ثبت شده در ایستگاه هواشناسی در روز i، \bar{x} و \bar{y} میانگین مقادیر ذکر شده و n تعداد مشاهدات می‌باشد.

۳- یافته‌ها و بحث

در پژوهش حاضر با به‌کارگیری مدل‌های درختی M5P، درخت گرادیان تقویتی (GBT)، جنگل تصادفی (RF) و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (RF-GA) دمای نقطه شب‌نم روزانه در دو ایستگاه اردبیل و پارس‌آباد با استفاده از پارامترهای هواشناسی به‌عنوان داده‌های ورودی در ۱۰ سناریو مختلف برآورد شد به‌منظور امکان‌سنجی استفاده از ترکیب‌های مختلف داده‌های ورودی براساس ضریب همبستگی بین پارامتر دمای نقطه شب‌نم هر ایستگاه با سایر پارامترهای هواشناسی ایستگاه در جدول (۳) آورده شده است.

جدول ۳- ضریب همبستگی بین پارامتر دمای نقطه شب‌نم هر ایستگاه با سایر پارامترهای هواشناسی

Table 3 Correlation coefficient statistic of DPT with other model input attributes in each station

Station	T	Tmin	Tmax	RH	S	W
Ardabil	0.874	0.886	0.812	0.092	0.314	-0.72
Parsabad	0.948	0.956	0.897	-0.398	0.369	0.077

لذا با استفاده از داده‌های هواشناسی حداقل، حداکثر و میانگین دما، رطوبت نسبی، ساعت آفتابی و سرعت باد، ۱۰ الگو از ترکیب‌های مختلف هواشناسی تعریف گردید (جدول ۴). سپس محاسبات واسنجی بر روی ۷۰٪ داده‌ها و

¹Root Mean Squared Error

²Willmott's Index of agreement

³Coefficient of Determination

مربعات خطای ۰/۲۹ بهترین پیش‌بینی را در بین روش‌های مذکور داشته است و عملکرد مدل RF-GA-6 در مرتبه دوم قرار گرفت. در این ایستگاه الگو شماره ۶ در همه روش‌ها بهترین نتایج را ارائه داد. مدل GBT-1 با ضریب تعیین

۰/۹۰۸، ضریب توافق ویلموت ۰/۷۸ و جذر میانگین مربعات خطای برابر با ۴/۳۸ ضعیف‌ترین نتایج را ارائه کرده و همچنین مانند ایستگاه اردبیل الگو شماره ۱ عملکرد ضعیفی را در بین سایر الگوها ارائه کرد.

جدول ۵- نتایج ارزیابی عملکرد مدل‌ها در مرحله صحت‌سنجی

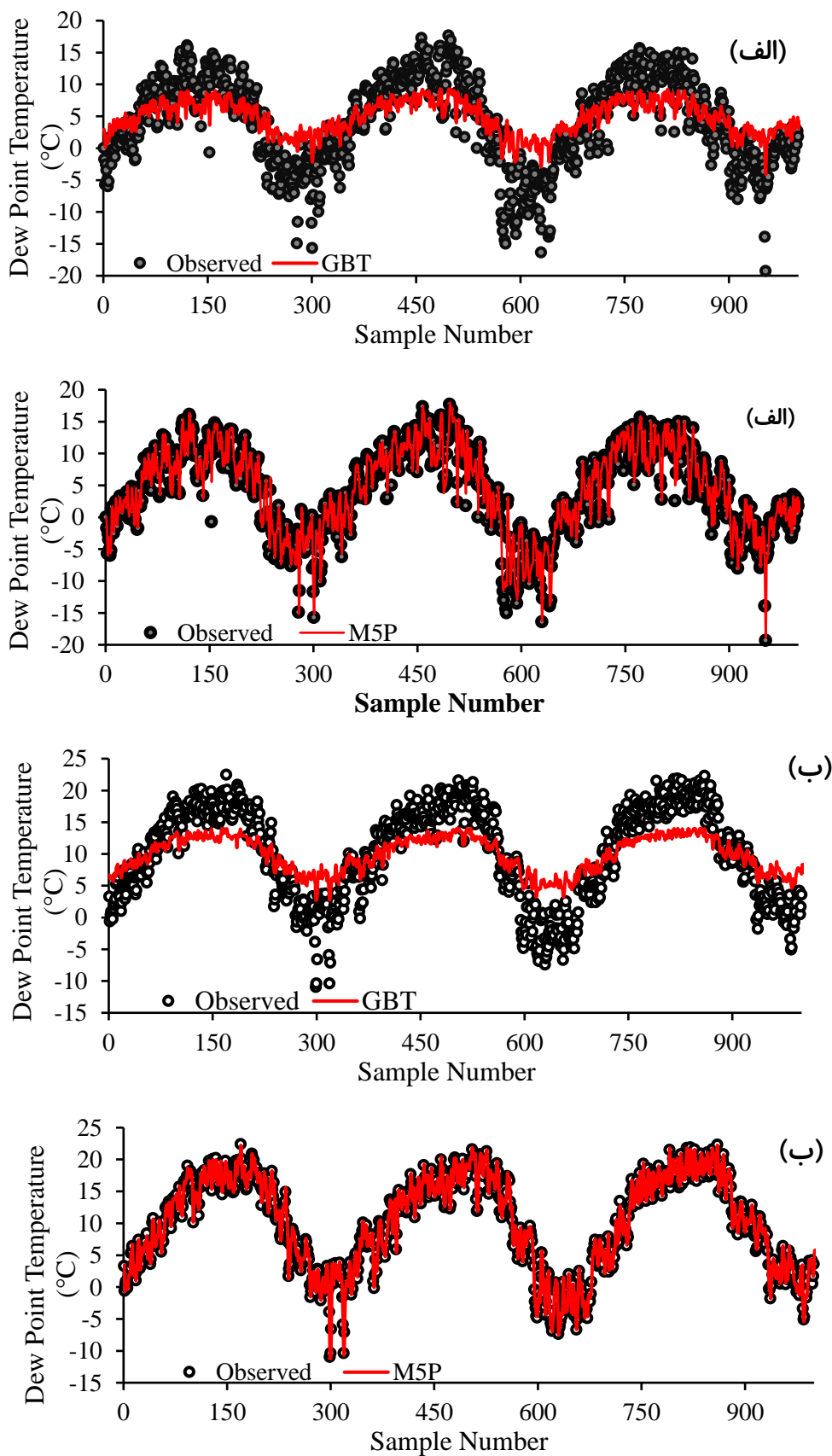
Table 5 Evaluation of the models performance in the testing phase

Station	Scenario	M5P			RF			RF- GA			GBT		
		R ²	WI	RMSE	R ²	WI	RMSE	R ²	WI	RMSE	R ²	WI	RMSE
Ardabil	1	0.715	0.913	3.73	0.706	0.910	3.79	0.715	0.913	3.72	0.713	0.722	4.72
	2	0.766	0.928	3.44	0.757	0.925	3.52	0.774	0.930	3.36	0.770	0.740	4.60
	3	0.773	0.931	3.37	0.771	0.930	3.41	0.769	0.930	3.36	0.769	0.740	4.60
	4	0.845	0.953	2.85	0.842	0.952	2.87	0.838	0.951	2.92	0.829	0.759	4.46
	5	0.789	0.937	3.25	0.787	0.935	3.29	0.781	0.934	3.33	0.777	0.743	4.58
	6	0.993	0.998	0.55	0.986	0.996	0.80	0.988	0.997	0.73	0.966	0.900	4.18
	7	0.851	0.955	2.80	0.846	0.953	2.82	0.846	0.953	2.84	0.831	0.760	4.45
	8	0.994	0.998	0.54	0.984	0.995	0.88	0.987	0.996	0.78	0.966	0.789	4.18
	9	0.994	0.998	0.55	0.984	0.995	0.88	0.987	0.996	0.79	0.966	0.789	4.18
	10	0.994	0.998	0.54	0.981	0.994	0.97	0.985	0.996	0.83	0.966	0.789	4.18
Parsabad	1	0.909	0.976	2.07	0.903	0.974	2.13	0.908	0.975	2.08	0.908	0.780	4.38
	2	0.920	0.979	1.95	0.917	0.978	1.99	0.916	0.978	1.99	0.920	0.786	4.34
	3	0.934	0.983	1.78	0.931	0.982	1.81	0.926	0.980	1.87	0.925	0.787	4.33
	4	0.939	0.982	1.82	0.934	0.980	1.89	0.927	0.980	1.87	0.929	0.789	4.32
	5	0.935	0.983	1.76	0.933	0.982	1.78	0.931	0.982	1.81	0.927	0.787	4.32
	6	0.999	1.000	0.29	0.992	0.998	0.61	0.995	0.999	0.49	0.964	0.901	4.26
	7	0.940	0.982	1.80	0.938	0.983	1.78	0.938	0.982	1.80	0.929	0.789	4.32
	8	0.999	1.000	0.29	0.990	0.998	0.67	0.994	0.998	0.55	0.964	0.794	4.26
	9	0.999	1.000	0.29	0.991	0.998	0.65	0.994	0.998	0.53	0.964	0.794	4.26
	10	0.999	1.000	0.29	0.989	0.997	0.73	0.992	0.998	0.59	0.964	0.794	4.26

ملاحظه‌ای افزایش پیدا می‌کند. نمودارهای تغییرات مقادیر دمای نقطه شبنم پیش‌بینی شده با استفاده از روش‌های مورد مطالعه با مقادیر حاصل از مشاهدات برای برترین و ضعیف‌ترین مدل‌ها در شکل (۶) قابل ملاحظه می‌باشد.

با توجه به نمودارهای تغییرات زمانی مقادیر دمای نقطه شبنم مشاهداتی و پیش‌بینی شده با استفاده از روش‌های مورد مطالعه و با به‌کارگیری پارامترهای هواشناسی به‌عنوان داده‌های ورودی، روند اشاره شده در مورد مدل‌های برتر قابل استنباط است. همانطور که اشاره شد دقت بالای مدل M5P برای هر دو ایستگاه اردبیل و پارس آباد قابل مشاهده است، چرا که انطباق نقاط برآورد شده از روش‌های مذکور بر نقاط حاصل از مشاهدات بیشتر است. عدم انطباق نقاط برآورد شده و حاصل از مشاهدات در مدل GBT نیز مشهود و نشان از عدم دقت این مدل در برآورد دمای نقطه شبنم می‌باشد. در شکل (۷) نیز نمودارهای پراکنش برای برترین و ضعیف‌ترین مدل‌ها قابل ملاحظه می‌باشد.

در حالت کلی، مدل درختی M5P در هر دو ایستگاه به‌عنوان مدل برتر در تخمین مقادیر دمای نقطه شبنم معرفی شد. مدل GBT با ارائه نتایج ضعیف به‌عنوان روشی ناکارآمد در پیش‌بینی مقادیر روزانه دمای نقطه شبنم در نظر گرفته شد. از طرفی، در ایستگاه پارس آباد روش‌های به‌کار گرفته شده با توجه به همبستگی بالای مقادیر دما و رطوبت نسبی با دمای نقطه شبنم دقت تخمین‌ها بالاتر بوده و عملکرد مطلوب‌تری (ضریب تعیین بالا و خطای کم) را نسبت به ایستگاه اردبیل داشتند و لازم به ذکر است که پارامتر دما شرط لازم ولی ناکافی در برآورد دمای نقطه شبنم بود. همانطور که در جدول (۳) ملاحظه می‌شود علی‌رغم اینکه مقدار همبستگی رطوبت نسبی با دمای نقطه شبنم پایین است اما احتمالاً پارامتر رطوبت نسبی در ترکیب خطی با پارامترهای دما، همبستگی بالایی با دمای نقطه شبنم داشته و با توجه به مقایسه نتایج ذکر شده، با اضافه شدن رطوبت نسبی به داده‌های ورودی، دقت مدل‌های مورد مطالعه به‌صورت قابل

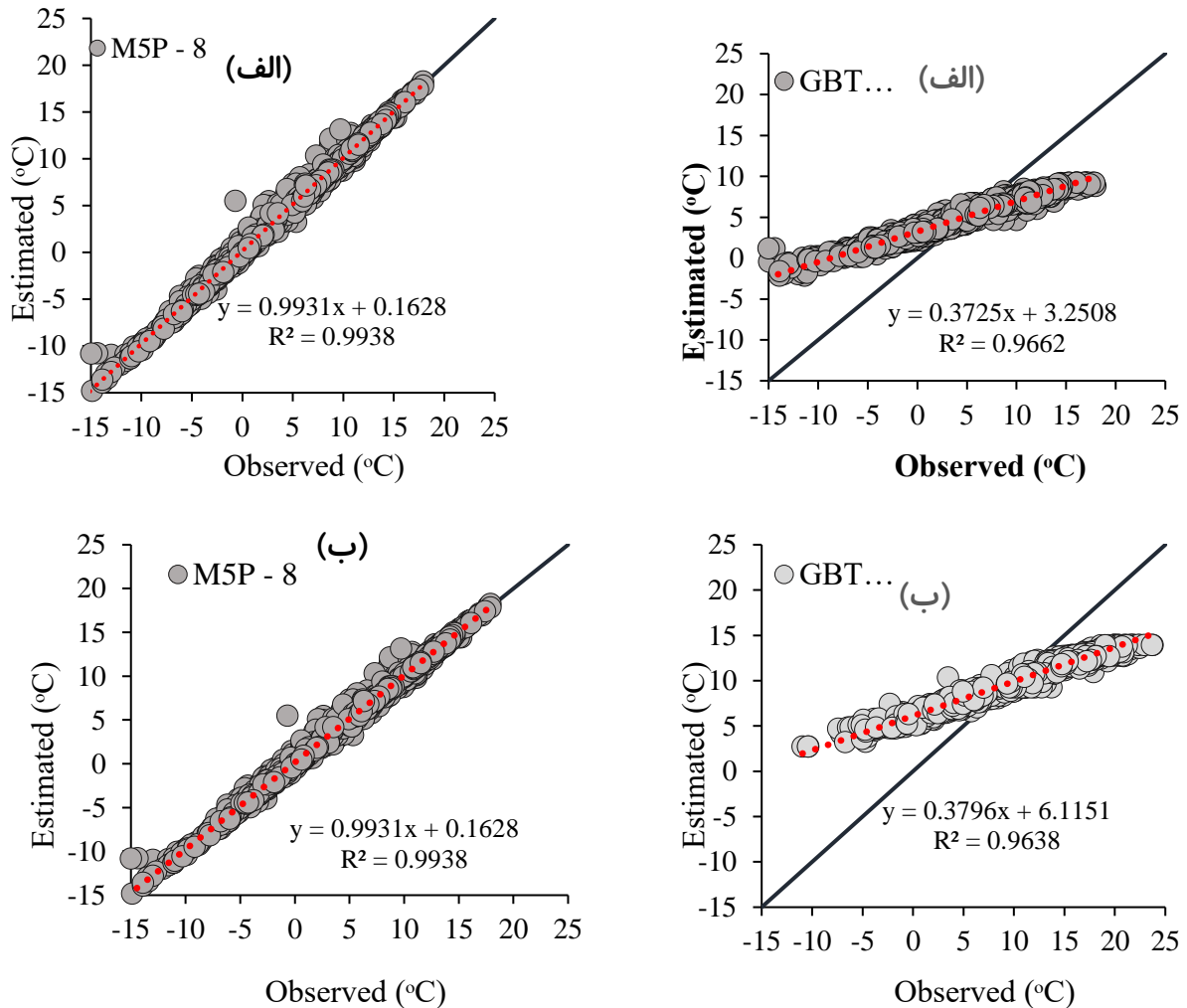


شکل ۶- نمودار تغییرات زمانی مقادیر دمای نقطه شبنم برای برترین و ضعیف‌ترین مدل‌ها در ایستگاه‌های: الف) اردبیل و ب) پارس‌آباد
 Fig. 6 Charts of temporal variations dew point temperature values for best and weakest models at: a) Ardabil and b) Parsabad station



مدل GBT در تحلیل فرآیند غیرخطی تخمین دمای نقطه شبنم باشد. اختلاف معنی‌داری بالایی بین عملکرد مدل‌های M5P و RF و GA-RF وجود ندارد، لذا با توجه به روابط ریاضی ارائه شده توسط مدل M5P، این مدل می‌تواند با اطمینان بالایی به‌عنوان یک مدل کاربردی برای تخمین دمای نقطه شبنم توصیه گردد.

در شکل (۷) نیز تخمین مقادیر نقاط حدی حداکثر و حداقل نیز دقت خوبی نشان داده است. برخلاف مدل GBT که پراکنش نقاط حول خط نیمساز زیاد بوده و همچنین اختلاف زیادی در تخمین مقادیر حدی حداکثر و حداقل دمای نقطه شبنم داشته که حاکی از توان پایین این مدل در برآورد این دما است. به نظر می‌رسد که این مورد ناشی از توانایی پایین



شکل ۷- نمودارهای پراکنش دمای نقطه شبنم در ایستگاه‌های: الف) اردبیل و ب) پارس آباد

Fig. 7 Scatter plots of dew point temperature at: a) Ardabil and b) Parsabad stations

تعریف شده دارد. علاوه بر این، مقایسه نتایج به‌دست آمده با نتایج حاصل از پژوهش (2019) Qasem et al. که نشان دادند با استفاده از پنج پارامتر شامل میانگین دمای هوا، رطوبت نسبی، فشار بخار واقعی، سرعت باد و ساعات آفتابی، مدل M5P با جذر میانگین مربعات خطای برابر با ۰/۳۷ به‌عنوان بهترین مدل در پیش‌بینی دمای نقطه شبنم است، مطابقت دارد.

در شکل (۸) مقدار جذر میانگین مربعات خطای هر یک از مدل‌های پیشنهادی به ازای هر الگوی تعریف شده در هر دو ایستگاه قابل مشاهده می‌باشد. همانطور که از شکل (۸) ملاحظه می‌شود، مدل‌های M5P، RF-GA و RF خطای کم و مدل GBT دارای خطای زیاد است و الگو شماره ۶ هم با پارامترهای دما و رطوبت نسبی کم‌ترین خطا را در بین الگوهای مختلف

دمای نقطه شبنم در دو ایستگاه هواشناسی اردبیل و پارس‌آباد برآورد شد. همچنین با به‌کارگیری الگوریتم ژنتیک مقادیر به‌دست‌آمده از مدل‌های RF بهینه‌سازی شد. در نهایت با مقایسه عملکرد مدل‌ها در سناریوهای مختلف در دو ایستگاه مورد مطالعه، نتایج زیر بدست آمد:

(۱) در ایستگاه اردبیل مدل M5P در ترکیب هشتم شامل داده‌های ورودی رطوبت نسبی، سرعت باد، حداقل، حداکثر و متوسط دما و در ایستگاه پارس‌آباد مدل مذکور در ترکیب ششم شامل داده‌های ورودی رطوبت نسبی، حداقل، حداکثر و متوسط دما مطلوب‌ترین نتایج را ارائه دادند.

(۲) در مقایسه نتایج بین مدل‌ها نیز به‌ترتیب مدل M5P، RF-GA، RF و GBT از دقت بیشتر به کمتر رتبه‌بندی شدند.

(۳) پارامترهای دما و رطوبت نسبی به‌عنوان متغیرهای غالب در پیش‌بینی دمای نقطه شبنم معرفی شدند.

(۴) با مقایسه نتایج حاصل از مدل‌های RF با RF-GA چنین استنباط شد که الگوریتم ژنتیک باعث بهبود عملکرد جنگل تصادفی شده و تأثیر مثبتی بر همه مدل‌ها داشته است.

دسترسی به داده‌ها

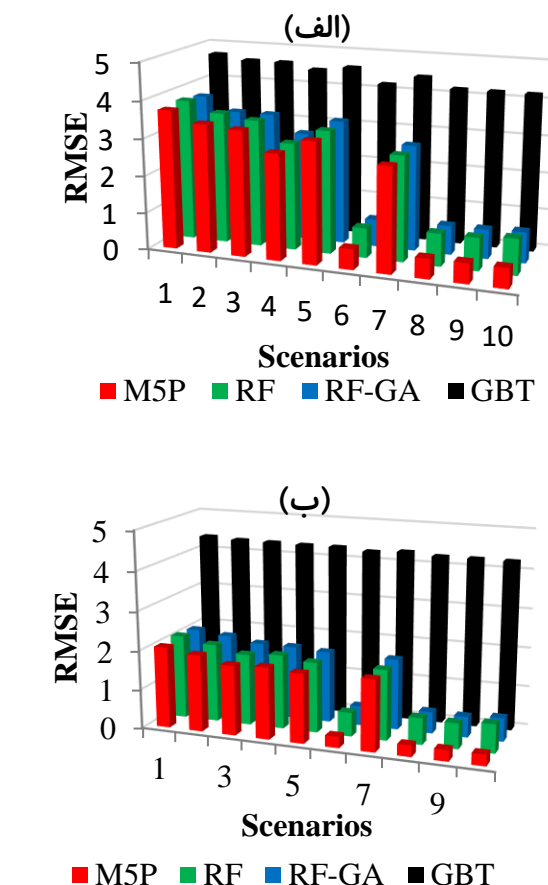
داده‌ها حسب درخواست، از طرف نویسنده مسئول از طریق ایمیل s.samadian@tabrizu.ac.ir قابل ارسال می‌باشد.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ گونه تضاد منافی در رابطه با نوشتن و یا انتشار این مقاله ندارند.

References

- Alizamir, M., Kim, S., Zounemat-Kermani, M., Heddad, S., Kim, N. W. and Singh, V. P. (2020). Kernel extreme learning machine: An efficient model for estimating daily dew point temperature using weather data. *Water*, 12(9), 2600. DOI: 10.3390/w12092600
- Amirmojahedi, M., Mohammadi, K., Shamshirband, S., Seyed Danesh, A., Mostafaeipour, A. and Kamsin, A. (2016). A hybrid computational intelligence method for predicting dew point temperature. *Environ. Earth. Sci.*, 75(5). DOI: 10.1007/s12665-015-5135-7



شکل ۸- مقادیر خطای مدل‌های پیشنهادی در ترکیب‌های مختلف داده ورودی در ایستگاه‌های: (الف) اردبیل و (ب) پارس‌آباد
Fig. 8 Error values of the proposed models per scenario at: a) Ardabil and b) Parsabad stations

۴- نتیجه‌گیری

در این پژوهش با استفاده از داده‌های مختلف هواشناسی در ترکیب‌های مختلف به‌عنوان ورودی مدل‌های درختی M5P، درخت گرادیان تقویتی (GBT)، جنگل تصادفی (RF) و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (RF-GA)،

- Breiman, L. (2001). *Random forest*. Kluwer Academic Publishers. Manufactured in The Netherlands. Springer Science and Business Media LLC. 5-32. DOI: 10.1023/a:1010933404324

- Deka, P. C., Patil, A. P., Yeswanth Kumar, P. and Naganna, S. R. (2017). Estimation of dew point temperature using SVM and ELM for humid and semi-arid regions of India. *ISH J. Hydraul. Eng.*, 24(2), 190-197. DOI: 10.1080/09715010.2017.1408037

- Dong, J., Wu, L., Liu, X., Li, Z., Gao, Y., Zhang, Y. and Yang, Q. (2020). Estimation of daily

- dew point temperature by using bat algorithm optimization based extreme learning machine. *Appl. Therm. Eng.*, 165, 114569. DOI: [10.1016/j.applthermaleng.2015.11.081](https://doi.org/10.1016/j.applthermaleng.2015.11.081)
- Mohammadi, K., Shamshirband, S., Petković, D., Yee, P. L. and Mansor, Z. (2016). Using ANFIS for selection of more relevant parameters to predict dew point temperature. *Appl. Therm. Eng.*, 96, 311–319. DOI: [10.1016/j.applthermaleng.2015.11.081](https://doi.org/10.1016/j.applthermaleng.2015.11.081)
- Naganna, S., Deka, P., Ghorbani, M., Biazar, S., Al-Ansari, N. and Yaseen, Z. (2019). Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water*, 11(4), 742. DOI: [10.3390/w11040742](https://doi.org/10.3390/w11040742)
- Qasem, S. N., Samadianfard, S., Sadri Nahand, H., Mosavi, A., Shamshirband, S. and Chau, K. w. (2019). Estimating daily dew point temperature using machine learning algorithms. *Water*, 11, 582. DOI: [10.3390/w11030582](https://doi.org/10.3390/w11030582)
- Samadianfard S., Salarifar M., Javidan S. and Mikaeili F. (2020). Estimation of daily reference evapotranspiration in humid climates using data-driven methods of gaussian process regression, support vector regression and random forest. *Environ. Water Eng.*, 6(4), 360–373. DOI: [10.22034/jewe.2020.241690.1394](https://doi.org/10.22034/jewe.2020.241690.1394) [In Persian]
- Samadianfard, S., Kargar, K., Shadkani, S., Hashemi, S., Abbaspour, A. and Safari, M. J. S. (2021). Hybrid models for suspended sediment prediction: optimized random forest and multi-layer perceptron through genetic algorithm and stochastic gradient descent methods. *Int. J. Sediment. Res.* 36(4), 512-523. DOI: [10.1007/s00521-021-06550-1](https://doi.org/10.1007/s00521-021-06550-1)
- Shank, D. B., McClendon, R. W., Paz, J. and Hoogenboom, G. (2008). Ensemble artificial neural networks for prediction of dew point temperature. *Appl. Artif. Intell.*, 22(6) 523-542. DOI: [10.1080/08839510802226785](https://doi.org/10.1080/08839510802226785)
- Shiri J., Kim S. and Kisi O. (2014). Estimation of daily dew point temperature using genetic programming and neural networks approaches. *Hydrol. Curr. Res.*, 45 (2): 165-181. DOI: [10.2166/nh.2013.229](https://doi.org/10.2166/nh.2013.229)
- Shiri, J. (2018). Prediction vs. estimation of dew point temperature: assessing GEP, MARS and RF models. *Hydrol. Curr. Res.*, 50(2), 633–643. DOI: [10.2166/nh.2018.104](https://doi.org/10.2166/nh.2018.104).
- Siasar, H., Honar T. and Abdolahipour, m. (2020). Comparing of generalized linear models, random forest and gradient boosting trees in estimation of reference crop evapotranspiration (Case Study: The Sistan Plain). *J. Water and Soil Sci.*, 23(4). DOI: [10.47176/jwss.23.4.4631](https://doi.org/10.47176/jwss.23.4.4631) [In Persian]
- Zhan, C., Gan, A. and Hadi, M. (2011). Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. *IEEE Trans Intell. Transp. Syst.*, 12 (4), 1549–1557. DOI: [10.1109/TITS.2011.2161634](https://doi.org/10.1109/TITS.2011.2161634)

How to cite this paper:

Osouli, M., Mikaeili, F. and Samadianfard, S. (2022). Evaluating capabilities of gradient boosted tree and optimized random forest models in estimating daily dew point temperature. *Environ. Water Eng.*, 8(3), 654–668. DOI: [10.22034/JEWE.2022.313219.1662](https://doi.org/10.22034/JEWE.2022.313219.1662)

