2023, Vol.10, No.3, Pages 87 to 98
Journal of Water and Sustainable Development

سال دهم، شماره ۳، ۱۴۰۲، صفحات ۸۷ تا ۹۸
نشریه آب و توسعه پایدار

Article Type: Applied Article

نوع مقاله: پژوهش کاربردی

Genetic Programming Method in Urban Water Consumption Prediction (Case Study: Najafabad City)

M. R. Alikhani[1], R. Moeini[2*]

1, 2- M.Sc. Student and Associate Professor, Department of Civil Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan, Iran.

*(Corresponding Author Email: r.moeini@eng.ui.ac.ir)

Received: 20-06-2023 Revised: 27-08-2023
Accepted: 05-09-2023 Available Online: 21-12-2023

روش برنامه‌نویسی ژنتیک (GP) در پیش‌بینی میزان مصرف آب شهری (مطالعه موردی: شهر نجف آباد)

محمدرضا علیخانی[1]، رامتین معینی[2*]

۱ و ۲- به‌ترتیب دانشجوی کارشناسی ارشد مهندسی و مدیریت منابع آب و دانشیار، گروه عمران، دانشکده مهندسی عمران حمل و نقل، دانشگاه اصفهان، اصفهان، ایران.

*(نویسنده‌ی مسئول، E-Mail: r.moeini@eng.ui.ac.ir)

تاریخ دریافت: ۱۴۰۲/۰۳/۳۰ تاریخ بازنگری: ۱۴۰۲/۰۶/۰۵
تاریخ پذیرش: ۱۴۰۲/۰۶/۱۴ تاریخ انتشار: ۱۴۰۲/۰۹/۳۰

Abstract

Improving the lifestyle of city residents is conditional on benefiting from high-quality urban infrastructure to satisfy daily demands. The urban water supply network is one of the most basic urban infrastructures, and its optimal design and service are essential during the planning period. Therefore, it is important to determine the actual amount of consumption and predict it for the future. For this purpose, in this research, a method based on artificial intelligence, i.e., genetic programming (GP), as well as Pearson's correlation coefficient data mining method, is proposed. The data mining method is applied here for the database, including daily data on temperature, precipitation, humidity, and the amount of daily water produced in Najafabad city (presenting the total water consumption) from the beginning of 2014 to the end of 2018, and the best set of input data vectors is selected. The selected data are used as input data vectors for the proposed. The obtained results are compared with the results of models based on artificial neural network (ANN). To investigate the performance of the models, R^2, RMSE, and NSE statistical indices are calculated. A comparison of the results indicates the acceptable performance of the proposed models based on the GP. In other words, the values of RMSE, NSE, R^2, and MAPE statistical indices for training data in the best GP model are equal to 3262.59 MCM, 0.80, 0.80, and 5.38%, respectively, and for test data equal to 3507.68 MCM, 0.78, 0.78, and 6.67%.

Keywords: Water Consumption, Prediction, Artificial Intelligence, Genetic Programming, Artificial Neural Network, Najafabad City.

چکیده

ارتقا سطح کیفی زندگی ساکنین شهرها به بهره‌مندی از زیرساخت شهری با کیفیت بالا، به‌منظور برآورده کردن نیازهای روزانه وابسته است. شبکه آبرسانی شهری یکی از اساسی‌ترین زیرساخت‌های شهری است که طراحی و سرویس‌دهی مطلوب آن در طول دوره طرح ضروری است. ازاین‌رو تعیین میزان واقعی مصرف و پیش‌بینی آن در آینده اهیمت دارد. به‌این‌منظور در این تحقیق، از یک روش بر مبنای هوش مصنوعی، روش برنامه‌نویسی ژنتیک (GP) و همچنین روش داده‌کاوی ضریب هم‌بستگی پیرسون استفاده شده است. روش‌های داده‌کاوی بر روی بانک داده شامل داده‌های روزانه دما، بارش، رطوبت و مقدار آب تولیدی روزانه برای شهر نجف آباد در اصفهان از ابتدای سال ۱۳۹۴ تا انتهای سال ۱۳۹۸اعمال و بهترین ترکیب از داده‌های ورودی انتخاب شده است. داده‌های منتخب به‌عنوان بردارهای ورودی برای مدل‌های پیشنهادی بر مبنای روش GP اعمال و مقدار آب تولیدی روزانه (نشان‌دهنده آب مصرفی کل)شهر نجف آباد پیش‌بینی شده است. نتایج به‌دست آمده با نتایج مدل‌هایی بر مبنای شبکه عصبی مصنوعی (ANN) مقایسه شده است. برای بررسی عملکرد مدل‌ها، شاخص‌های آماری R^2، RMSE و NSE محاسبه شده است. مقایسه نتایج نشان‌دهنده عملکرد قابل قبول مدل‌های پیشنهادی بر مبنای GP می‌باشد. به‌عبارت‌دیگر، مقدار شاخص‌های آماری RMSE، NSE، R^2 و MAPE برای داده‌های آموزش در بهترین مدل GP به‌ترتیب برابر با ۳۲۶۲/۵۹ MCM، ۰/۸۰، ۰/۸۰ و ۵/۳۸ درصد و برای داده‌های آزمایش به ترتیب برابر با ۳۵۰۷/۶۸ MCM، ۷۸/۷۸، ۰/۰ و ۶/۶۷ درصد می‌باشد.

واژه‌های کلیدی: مصرف آب، پیش‌بینی، هوش مصنوعی، برنامه‌نویسی ژنتیک، شبکه عصبی مصنوعی، شهر نجف‌آباد.

How to cite this article: Alikhani, M. R., & Moeini, R. (2023). Genetic Programming Method in Urban Water Consumption Prediction (Case Study: Najafabad City). Journal of Water and Sustainable Development, 10(3): 87-98. doi: http://dx.doi.org/10.22067/jwsd.v10i3.2306-1251

مقدمه

پیش‌بینی مقدار تقاضای آب یکی از چالش‌های مهم در مدیریت منابع آب است. آب یک منبع حیاتی و محدود برای انسان‌ها است که برای نیازهای شخصی، خانگی، صنعتی، کشاورزی و محیط‌زیستی استفاده می‌شود. با رشد جمعیت، توسعه اقتصادی، تغییرات اقلیم و تغییر الگوهای مصرف، تقاضای آب در حال افزایش است. درحالی‌که منابع آب در بسیاری از مناطق دنیا با کمبود، آلودگی و نابرابری مصارف مواجه هستند. بنابراین، پیش‌بینی مقدار تقاضای آب می‌تواند به بهبود برنامه‌ریزی و تخصیص منابع آب کمک کند. به‌عبارت‌دیگر، با انجام یک پیش‌بینی دقیق و قابل اعتماد، مسئولین می‌توانند تصمیمات بهتر و به هنگام اتخاذ نمایند. ازجمله این تصمیمات می‌توان به اجرای پروژه‌های ساخت‌وساز جدید، اصلاح قوانین و مقررات، تعرفه‌گذاری و حمایت‌های مالی، اجرای برنامه‌های صرفه‌جویی و کاهش هدر رفت آب اشاره نمود. علاوه‌براین، پیش‌بینی مقدار تقاضای آب نه تنها به بهبود کارآمدی و پایداری سامانه‌های آبرسانی شهری کمک می‌کند، بلکه می‌تواند از بروز بحران‌ها و فاجعه‌های اجتماعی جلوگیری نماید. به‌عبارت‌دیگر، کمبود آب چالش جدی برای امنیت غذایی، سلامت عمومی، توسعه انسانی و ثبات منطقه‌ای و جهانی است. ازجمله پیامدهای کمبود آب می‌توان به کاهش محصولات کشاورزی، گسترش بیماری‌ها، افزایش فقر، مهاجرت، تنش‌ها و تعارض‌ها اشاره نمود. لذا، با پیش‌بینی مقدار تقاضای آب، می‌توان از رخ دادن این پیامدها جلوگیری و به حفظ منابع آب برای نسل‌های آینده کمک کرد. یکی از مولفه‌های آب مصرفی، میزان آب شهری است که پیش‌بینی میزان تقاضای آب در این بخش لازم و ضروری است.در حالت کلی، روش‌های مختلفی به‌منظور پیش‌بینی مقدار تقاضای آب در مناطق شهری و مدیریت مصرف پیشنهاد شده است که هریک مزایا، محدودیت‌ها و فرضیات متفاوت دارند. امروزه استفاده از روش‌های هوش مصنوعی به دلیل قابلیت‌های آن بیشتر مورد توجه محققین می‌باشد. در ادامه به برخی از تحقیقات انجامشده در این حوزه اشاره می‌شود.

Austin و Richard (۲۰۱۰) مدل‌هایی براساس رگرسیون به‌منظور پیش‌بینی تقاضای آب شهری در یک خانوار در جامعه آماری با اندازه‌گیری هر دو ماه یکبار در شهر سیاتل در کشور آمریکا پیشنهاد نمودند. به‌این‌منظور سه روش رگرسیونی پیشنهاد شد. بررسی نتایج نشان داد عملکرد مدل‌ها در تخمین مقادیر مصرف در منطقه مورد مطالعه مناسب است.

Lee و همکاران (۲۰۱۱) تاثیرات طرح‌های تشویقی دولت شامل طرح تخفیف و تعویض ماشین لباسشویی، توالت و سردوش حمام به ساکنین میامی را در طول ۴ سال بررسی نمودند. بررسی نتایج نشان داد استفاده از سردوش‌ها، سرویس‌های بهداشتی و ماشین لباس‌شویی با مصرف بهینه آب، تاثیر زیادی در کاهش تقاضای آب خانگی داشته است.

Suero Francisco و همکاران (۲۰۱۲) مقادیر مصرف، مدت زمان مصرف و زمان مصرف را به‌صورت جداگانه برای هر وسیله مصرف‌کننده آب، در مدت زمان ۲ هفته پیش از تعویض و ۴ هفته پس از تعویض بررسی نمودند. در این پژوهش برای هر یک از وسایل مصرف‌کننده آب، مقادیر ذخیره مشاهده شده با مقادیر ذخیره پیش‌بینی شده به‌وسیله مدل‌های تحلیل ساده، رگرسون و مدل‌های تحلیل ترکیبی مقایسه شد.بررسی نتایج بیانگر تاثیر زیادوسایل کاهنده مصرف (کاهش دبی جریان) و پارامترهای رفتاری مصرف‌کنندگان (طول زمان استفاده و فرکانس استفاده) بود.

Romano و همکاران (۲۰۱۴) عوامل تاثیرگذار بر تقاضای آب در شهرهای بزرگ هر استان ایتالیا را با استفاده از مدل اثر ترکیبی خطی و روش حداکثر احتمال محدود تعیین نمودند. نتایج نشان‌دهنده تاثیر کاهنده تعرفه‌های اعمال شده بر مصرف آب بود. همچنین مقدار مصرف با افزایش سطح درآمد هر خانوار یا افزایش جمعیت در شهرهای بزرگ افزایش یافت.

Sebri (۲۰۱۶) با مطالعه بیش از ۶۰۰ مورد پیش‌بینی مقدار تقاضای آب شهری و تحلیل متارگرسیون[۱] نشان داد دقت پیش‌بینی در هر روش، به‌طور چشمگیری به نوع مطالعات ازجمله بازه زمانی تقاضا، روش مدل‌سازی، اندازه نمونه، مشخصات مدل و افق زمانی پیش‌بینی بستگی دارد.

Brentan و همکاران (۲۰۱۷) از مدل رگرسیون بردار پشتیبان (SVR)[۲]، به‌عنوان یکی از بهترین روش‌های یادگیری ماشین، برای پیش‌بینی کوتاه مدت تقاضای آب شهری استفاده نمودند. در مدل پیشنهادی از یک سری زمانی فوریه[۳] برای بهبود عملکرد مدل پیش‌بینی تقاضا شهر فرانکا[۴] در کشور برزیل استفاده شد. نتایج نشان داد با استفاده از سری زمانی فوریه، خطاها و آشفتگی‌های ذاتی مدل رگرسیون ثابت[۵] حذف شد که این‌کار باعث بهبود عملکرد مدل شد.

Shabaniو همکاران (۲۰۱۸) یک مدل ترکیبی بر مبنای الگوریتم برنامه‌نویسی بیانی ژنی (GEP) و یادگیری بدون ناظر[۶] را با هدف پیش‌بینی مقدار تقاضای آب شهر میلان در کشور ایتالیا در کوتاه‌مدت توسعه دادند. نتایج تحقیق نشان داد زمانی‌که الگوریتم GEP با روش‌های یادگیری بدون نظارت ترکیب شود، خوشه‌بندی داده‌های ورودی برای افزایش دقت نتایج مدل الزامی است.

Pacchin و همکاران (۲۰۱۹) کارایی و میزان دقت شِش مدل مختلف ازجمله مدل بر پایه شبکه عصبی، مدل مبتنی بر الگو[۷]، دو مدل مبتنی بر تکنیک حرکت پنجره[۸]، مدل زنجیره‌ای احتمالی مارکوف[۹] و مدل معیار ساده[۱۰] برای پیش‌بینی کوتاه‌مدت مقدار تقاضای آب شهری در شهر پادوآ[۱۱] در کشور ایتالیا را مقایسه نمودند. نتایج نشان‌دهنده بهترین عملکرد از مدل‌های مبتنی بر تکنیک حرکت پنجره، بود.

Xenochristou و Kapelan (۲۰۲۰) برای مقایسه و پیش‌بینی دقیق تقاضای آب در مقیاس‌های مختلف از پارامترهایی همچون مقدار مصرف در سطح هر خانوار، مکان سکونت خانوار برای ۱۲۰ خانوار در شهر اِکستِر[۱۲] در کشور انگلستان و داده‌های هواشناسی به‌عنوان ورودی‌های مدل ماشین تقویت‌کننده گرادیان (GBM)[۱۳] استفاده کردند. نتایج تحقیق نشان داد هرچه اندازه جامعه مورد بررسی کوچکتر شود دقت نتایج به‌دست آمده از مدل، به‌صورت نمایی کاهش و تاثیر پارامترهای مقدار مصرف در گذشته و ویژگی‌های هر خانوارِ مصرف‌کننده نسبت به پارامترهای هواشناسی افزایش می‌یابد. Smolak و همکاران (۲۰۲۰) برای پیش‌بینی کوتاه‌مدت مقدار مصرف آب از داده‌های مصرف آب در ۵۱ روز گذشته و همچنین داده‌های ثبت شده از جابه‌جایی ساکنین شهری در این بازه زمانی استفاده نمودند. در این تحقیق، برای مقایسه دقت نتایج از مدل کلاسیک و الگوریتم‌های یادگیری ماشین شامل ARIMA[۱۴]، SVR[۱۵]، RF[۱۶] و ERF[۱۷] استفاده شد. نتایج این تحقیق نشان داد استفاده از داده‌های ثبت شده مربوط به جابه‌جایی جمعیت شامل جمعیت ساکن و شناور بهبود نتایج را به‌همراه خواهد داشت.

Shuang و Zhao (۲۰۲۱) با استفاده از داده‌های مربوط به وضعیت اقتصادی، شرایط اجتماعی ساکنین، منابع آب در دسترس و مقدار آب مصرفی عملکرد یازده مدل یادگیری ماشین را برای پیش‌بینی مقدار آب مصرفی در مناطق پِکن، تیانجین و هبی[۱۸] در کشور چین، بررسی و مقایسه نمودند. بررسی نتایج نشان داد با استفاده از روش GBDT[۱۹]، نتایج دقیق‌تری در مقایسه با روش‌های یادگیری ماشین به‌دست می‌آید.

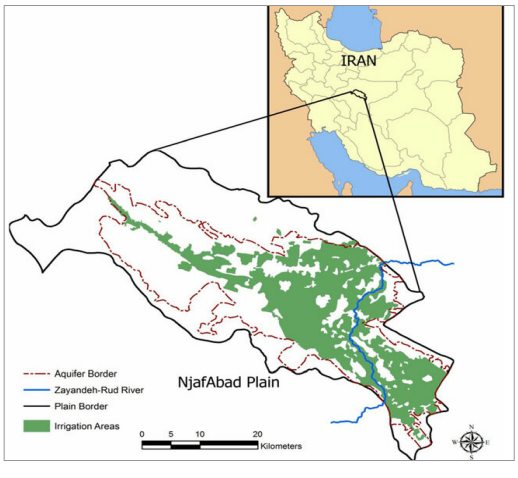
Kühnert و همکاران (۲۰۲۱) به‌منظور بهبود فرآیند تصمیم‌گیریِ تامین آب شرب شهری، عملکرد روش شبکه حافظه کوتاه-بلندمدت (LSTM)[۲۰] را با سایر روش‌های پیش‌بینی مقدار تقاضای آب مصرفی شهری و با استفاده از داده‌های شبکه تامین آب در یک روستای کوچک در کشور آلمان به‌عنوان ورودی مدل پیش‌بینی، مقایسه نمودند. در این تحقیق، بهترین نتایج با استفاده از روش LSTM به‌دست آمد.

بررسی سوابق تحقیق، نشان‌دهنده عملکرد مناسب استفاده از روش‌های هوش مصنوعی به‌منظور پیش‌بینی میزان مصرف و تقاضای آب شهری می‌باشد. مدل برنامه‌نویسی ژنتیک (GP) یکی از روش‌های هوش مصنوعی است که در حوزه‌های مختلف از آن استفاده شده است و عملکرد آن مناسب بوده است (Kazemi و Barati، ۲۰۲۲؛ Barati و همکاران، ۲۰۱۴). مهمترین قابلیت آن ارائه یک رابطه صریح برای میزان پارامتر خروجی بر مبنای پارامترهای ورودی تاثیرگذار می‌باشد. بنابراین در این تحقیق، این روش در تعیین میزان آب مصرفی شهر نجف‌آباد استفاده شده و عملکرد آن با مدل شبکه عصبی مصنوعی (ANN)[۲۱] مقایسه شده است. بررسی نتایج نشان‌دهنده عملکرد خوب روش پیشنهادی می‌باشد.

منطقه مطالعاتی

در این تحقیق، شهر نجف‌آباد واقع در استان اصفهان به‌عنوان مطالعه موردی انتخاب شده است. در این بخش اطلاعات مربوط به شهر نجف‌آباد، مرکز شهرستان نجف‌آباد شامل اطلاعات جغرافیایی، هواشناسی و مقدار آب تولیدی روزانه ارائه شده است.

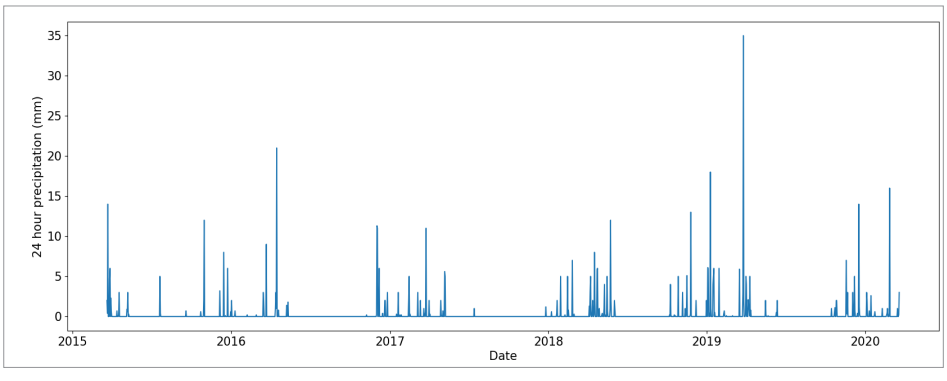
شهر نجف‌آباد در ۵۱ درجه و ۲۴ دقیقه طول شرقی از نصف‌النهار گرینویچ و ۳۸ درجه و ۳۲ دقیقه عرض شمالی از استوا قرار دارد. این منطقه از ارتفاعات دالان کوه با تراز ارتفاعی ۳۴۵۰ متر واقع در جنوب غربی حوضه که در ۱۰/۲۵ کیلومتری جنوب شرقی دامنه قرار دارد و در جهت شرق به طول ۸۱ کیلومتر تا فاصله ۲۱/۵ کیلومتری غرب شهر اصفهان قرار دارد. در شکل (۱) موقعیت تقریبی این شهر در استان اصفهان و کشور ایران نشان داده شده است.



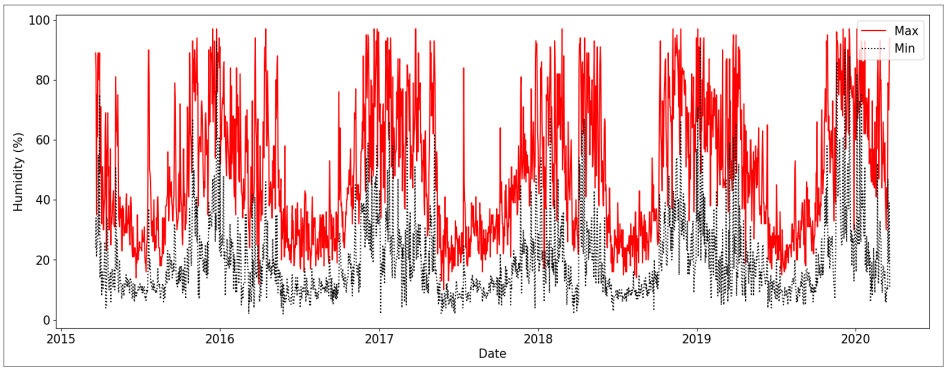
شکل ۱- موقعیت شهرستان نجف‌آباد در استان اصفهان (Goodarzi و همکاران، ۲۰۱۶)

در این تحقیق به‌منظور پیش‌بینی مقدار نیاز آب شهری شهرستان نجف‌آباد، از دو دسته داده روزانه شامل داده‌های هواشناسی و مقدار آب تولیدی ورودی به شهر، از ابتدای سال ۱۳۹۴ تا پایان سال ۱۳۹۸، به‌عنوان ورودی مدل‌های پیش‌بینی استفاده شده است که جزئیات آنها در ادامه ارائه شده است.به دلیل ماهیت سری زمانی مساله مورد بررسی و در دسترس بودن داده‌های مقدار آب تولیدی از ابتدای سال ۱۳۹۴ تا انتهای سال ۱۳۹۸، برای بررسی تاثیر تاخیرهای زمانی که از یک روز تا ۱ سال (۳۶۵ روز) می‌باشند، داده‌های مربوط به سال اول (۱ فروردین ۱۳۹۴ تا پایان ۲۹ اسفند ۱۳۹۴) از مجموعه داده‌های آموزش و آزمایش حذف شد و داده‌های باقی‌مانده به دو گروه آموزش و آزمایش و صحت‌سنجی تقسیم شدند. داده‌های آموزش مساله ۷۰ درصد ابتدایی مجموعه داده‌ها (۱۰۲۲ روز) و داده‌های آزمایش و صحت‌سنجی، ۳۰ درصد باقی‌مانده مجموعه داده‌ها (۴۳۹ روز)

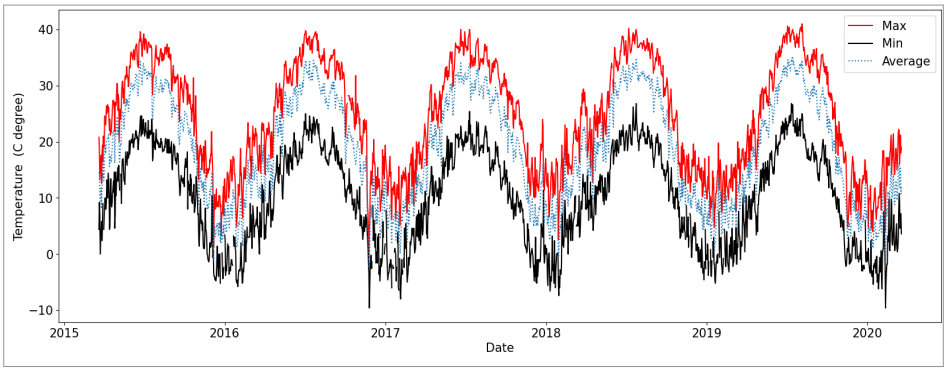
می‌باشند. در مجموع، داده‌های آموزش شامل بازه زمانی از ۱ فروردین ۱۳۹۵ تا پایان روز ۱۵ دی ۱۳۹۷ و داده‌های آزمایش بازه زمانی از ۱۶ دی ۱۳۹۷ تا ۲۹ اسفند ۱۳۹۸ را شامل می‌شوند. باتوجه‌به رژیم کم‌باران و نبود رودخانه دائم در این منطقه، تامین آب مورد نیاز ساکنین از محل چاه‌ها و قنات‌ها می‌باشد. اما در سال‌های اخیر، تامین آب شهر نجف‌آباد از طریق شهر اصفهان انجام می‌شود. در این تحقیق از داده‌های مربوط به مقدار آب تولیدی روزانه (میزان آب ورودی شهر که نشان‌دهنده مصارف کل روزانه شهری است) که توسط اداره آب و فاضلاب استان اصفهان ارائه شده است، استفاده شده است.علاوه‌براین، از داده‌های ایستگاه هواشناسی سینوپتیک استان اصفهان ازجمله دما، بارش و رطوبت نیز استفاده شده است. اطلاعات هواشناسی مورد استفاده ازجمله داده‌های مربوط به بارش، دما و رطوبت به ترتیب در شکل (۲)، (۳) و (۴) ارائه شده است.



شکل ۲- میزان بارش روزانه شهر نجف‌آباد در بازه زمانی مورد مطالعه



شکل ۳-رطوبت نسبی شهر نجف‌آباد در بازه زمانی مورد مطالعه



شکل ۴-دمای شهر نجفآباد در بازه زمانی مورد مطالعه

رطوبت نسبی، دما و بارش روزانه سه پارامتر اصلی هستند که بر وضعیت اقلیمی شهرستان نجف آباد تأثیر می‌گذارند. بررسی داده‌های تاریخی ارائه شده در شکل‌های (۲)، (۳) و (۴) در بازه زمانی ۵ ساله نشان دهنده این است میانگین بارش سالانه در این شهرستان ۲۷۶ میلیمتر است که کمتر از میانگین کشوری ۲۵۰ میلیمتر می‌باشد. علاوه‌براین، میانگین دمای سالانه در نجف‌آباد

۱۶/۵ درجه سانتی‌گراد است که بیشتر از میانگین کشوری ۱۵/۲ درجه سانتی‌گراد می‌باشد. همچنین، میانگین رطوبت نسبی سالانه در این شهرستان ۳۵ درصد است که کمتر از میانگین کشوری ۴۰ درصد می‌باشد. بررسی نتایج نشان‌دهنده رابطه معکوس دما با رطوبت نسبی و بارش روزانه است. علاوه‌براین، آمار نشان دهنده این است که شهرستان نجف آباد در یک اقلیم خشک و گرم قرار دارد.

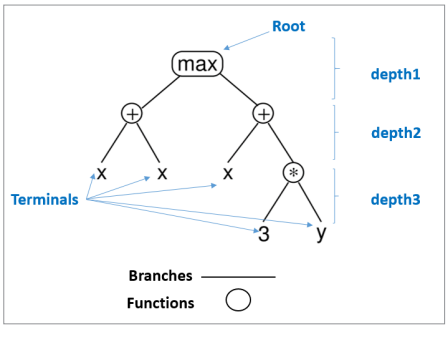
مواد و روش‌ها

در این تحقیق، ابتدا با استفاده از داده‌های روزانه مقدار آب تولیدی و داده‌های هواشناسی و استفاده از روش برنامه‌نویسی ژنتیک، در طی ۵ سال، مقدار آب تولیدی روزانه که نشان‌دهنده کل آب مصرفی روزانه در شهر نجف‌آباد است، پیش‌بینی شده است.نتایج این روش با نتایج به‌دست آمده از مدل شبکه عصبی مصنوعی مقایسه شده است. در این بخش توضیحاتی در خصوص روش‌های استفاده شده ارائه می‌شود.

۱-برنامه نویسی ژنتیک (GP)

روش برنامه‌نویسی ژنتیک یکی از تکنیک‌های یادگیری ماشین است که بدون نیاز به تعریف دقیق ساختار جوابِ یک مساله توسط کاربر، جواب مساله بر مبنای شبیه‌سازی فرآیند تکامل تعیین می‌شود. در این روش، فرآیند یافتن جواب مسائل به این صورت است که در گام نخست جمعیتی از جواب‌های مساله به‌صورت تصادفی در غالب برنامه‌های کامپیوتری ایجاد می‌شوند. در روش برنامه‌نویسی ژنتیک با اجرای هر یک از این برنامه‌های کامپیوتری و مقایسه خروجی این برنامه‌ها (داده‌های محاسباتی) با داده‌های مشاهداتی، برازندگی هر برنامه کامپیوتری ارزیابی می‌شود. نتایج سنجش برازندگی هر برنامه کامپیوتری به‌صورت کمی و در غالب اعداد با یکدیگر مقایسه می‌شود. برنامه‌های کامپیوتری که در مرحله سنجش برازندگی، نسبت به سایر برنامه‌های هم‌نسل خود جمعیت برتری دارند برای تولید مثل انتخاب می‌شوند و براین‌اساس نسل بعدی ایجاد می‌شود. عملگرهای ژنتیکی که برای ایجاد نسل جدید برنامه‌های کامپیوتری از نسل فعلی استفاده می‌شوند، جهش و آمیزش هستند که در ادامه به تفصیل بررسی می‌شوند (Silva و Almeida، ۲۰۰۳). ذکر این نکته ضروری است که برنامه‌ریزی ژنتیک شاخه‌ای از الگوریتم ژنتیک (GA)[۲۲] است واز قوانین تکاملی موجود در طبیعت الهام گرفته است. تفاوت اصلی این روش با الگوریتم ژنتیک در پردازش مساله و نحوه نمایش پاسخ‌ها است. در الگوریتم ژنتیک یک رشته از اعداد به‌عنوان پاسخ نمایش داده می‌شود، در صورتی‌که در برنامه‌نویسی ژنتیک، برنامه‌های کامپیوتری به‌عنوان پاسخ ارائه می‌شود. به‌عبارت‌دیگر، الگوریتم ژنتیک برای حل مسائل بهینه‌سازی گسسته و غیرخطی مناسب است، درحالی‌که برنامه‌نویسی ژنتیک برای شبیه‌سازی و حل مسائل پیچیده‌تر که نیاز به تولید برنامه‌های کامپیوتری دارد، مناسب‌تر می‌باشد.

در روش برنامه‌نویسی ژنتیک افراد جمعیت (جواب‌ها) به شکل برنامه‌های کامپیوتری (روابط رگرسیونی متغیرهای مستغل و متغیر وابسته) ارائه می‌شوند که به‌صورت درختی[۲۳] نمایش داده می‌شوند. به‌عنوان نمونه یکی از افراد جمعت به شکل عبارت Max[(x+x),(x+3×y)] ارائه شده که در شکل (۵) نحوه نمایش و اجزا تشکیل‌دهنده آن به شکل یک درخت در روش برنامه‌نویسی ژنتیک نمایش شده است. متغیرها (x ,y) و اعداد ثابت (عدد ۳) در این برنامه کامپیوتری را می‌توان به برگ‌های این درخت تشبیه کرد که در روش برنامه‌نویسی ژنتیک پایانه[۲۴] نامیده می‌شود. عملگرهای ریاضی (+ و × و Max) در این برنامه کامپیوتری گره‌های درونی این درخت هستند که در روش برنامه‌نویسی ژنتیک توابع[۲۵] نامیده می‌شوند. همچنین مفاهیمی از قبیل ریشه، عمق، شاخه و پایانه در این شکل نشان داده شده است.



شکل ۵- نحوه نمایش مساله نمونه در روش برنامه‌نویسی ژنتیک

گام اول در حل مسئله به‌وسیله برنامه‌نویسی ژنتیک، ایجاد جمعیت اولیه است. همانند سایر تکنیک‌های محاسبات تکاملی، در روش برنامه‌نویسی ژنتیک، جمعیت اولیه به‌صورت تصادفی ایجاد می‌شود. برای ایجاد جمعیت اولیه در روش برنامه‌نویسی ژنتیک از سه روش Grow، Full و Ramped half and half می‌توان استفاده نمود.

در مرحله بعد پس از ایجاد یک نسل از جمعیت، با اجرای هرکدام از برنامه‌های تولیدی (اعضا جمعیت) و مقایسه داده‌های خروجی آن‌ها با داده‌های مشاهداتی، مقدار برازندگی هر عضو به‌صورت کمّی تعیین می‌شود. پس از ایجاد جمعیت اولیه، به‌منظور ایجاد نسل بعدی از جمعیت، همانند سایر روش‌های محاسبات تکاملی فرآیند انتخاب انجام می‌شود. در فرآیند انتخاب جواب‌ها با برازندگی بالا، به‌عنوان والدین انتخاب و سپس عملگرهای ژنتیکی بر روی آنها اعمال می‌شوند تا نسل جدیدی از جواب‌ها تولید شود. این روند تا زمانی‌که شرط توقف روش برآورده شود، ادامه می‌یابد.

در روش برنامه‌نویسی ژنتیک پنج روش متفاوت برای انتخاب والدین پیشنهاد شده است. در فرآیند انتخاب، جواب‌ها با برازندگی بالا، به‌عنوان والدین انتخاب و سپس عملگرهای ژنتیکی بر روی آنها اعمال می‌شوند تا نسل جدیدی از جواب‌ها تولید شود. این روند تا زمان رسیدن به شرط توقف ادامه می‌یابد. در روش برنامه‌نویسی ژنتیک پنج روش متفاوت برای انتخاب والدین ازجمله روش چرخ‌گردان (roulette)، روش sus، روش tournament، روش lexictour و روش doubletour پیشنهاد شده است. به‌عنوان‌نمونه، در روش چرخ‌گردان اندازه سهم اختصاصی به هر یک از اعضای جمعیت متناسب با مقدار برازندگی آن‌ها می‌باشد. در ادامه، با چرخش چرخ‌گردان هر کدام از اعضای جمعیت که در مقابل نقطه انتخاب قرار بگیرد، برای تولید مثل انتخاب می‌شود. به‌عنوان‌نمونه، در روش sus، فرآیند انتخاب براساس چرخش یک چرخ‌گردان که سهمی برابر به هر یک از اعضای جمعیت اختصاص داده، انجام می‌شود. در روش برنامه‌نویسی ژنتیک پس از انتخاب والدین، با هدف ایجاد نسل جدید عملگرهای ژنتیکی بر روی آن‌ها اعمال می‌شود. با انتخاب تصادفی بخشی از یک والد (Parent) و تغییر آن بخش، یک جواب جدید (Child) تولید می‌شود. این فرآیند جهش نامیده می‌شود. بنابراین در فرآیند جهش بخشی از یک جواب به‌صورت تصادفی تغییر می‌یابد تا جواب جدیدی تولید شود و الزاماً نیازی به وجود ۲ والد نمی‌باشد. علاوه‌براین، در فرآیند اعمال عملگر ژنتیکی آمیزش، دو جواب انتخاب شده و سپس به‌صورت تصادفی قسمتی از هر یک از آن‌ها با یکدیگر ترکیب می‌شوند تا جواب جدیدی تولید شود. بنابراین در فرآیند آمیزش، با حضور دو جواب به‌عنوان والدین، یک جواب جدید به‌عنوان فرزند ایجاد می‌شود. از دیگر پارامترهای روش GP برای تعیین بهترین جواب مساله، تعیین تعداد فرزند مورد انتظار برای هر والد می‌باشد. این فرآیند در روش برنامه‌نویسی ژنتیک با استفاده از سه روش قدر مطلق، Rank 85 و روش Rank 89 انجام می‌شود.

۲- مدل شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی زیرمجموعه‌ای از علم هوش مصنوعی هستند که براساس ساختار و عملکرد مغز انسان طراحی شده‌اند. این سیستم، یک سیستم داده‌پردازی اطلاعات است که با هدف شبیه‌سازی شبکه‌های عصبی طبیعی معرفی شده است.این شبکه‌ها توانایی پردازش داده‌های پیچیده و غیرخطی را دارند و در بسیاری از کاربردهای یادگیری ماشین، ازجمله تشخیص الگو، تجزیه و تحلیل داده‌ها، تحلیل و پیش‌بینی سری‌های زمانی و موارد فراوان دیگر استفاده می‌شوند.این شبکه‌ها با تحلیل و بررسی داده‌های ثبتشده از گذشته، قابلیت یادگیری روابط و الگوهای موجود در این داده‌ها را دارند. در زمینه پیش‌بینی آب مصرفی و نیاز آبی، شبکه‌های عصبی مصنوعی با بررسی و تحلیل داده‌های جمع‌آوری شده در گذشته به‌عنوان داده‌های آموزش، قابلیت یادگیری الگوهای مرتبط با تقاضای آب را دارند و می‌توانند با استفاده از این الگوهای کشفشده، پیش‌بینی دقیق‌تری در آینده ارائه دهند.

در حالت کلی نورون‌ها[۲۶] واحد سازنده سیستم‌های عصبی هستند که پردازش اطلاعات را انجام می‌دهند. عملکرد یک سیستم عصبی، در نحوه ارتباط نورون‌ها با یکدیگر تعریف می‌شود. دسته‌بندی نورون‌ها براساس تعداد بردار داده‌های ورودی امکان‌پذیر است. همچنین، توابع تبدیل و انتقال از دیگر پارامترهای مهم در مدل شبکه عصبی مصنوعی می‌باشند که براساس آن خروجی مدل تعیین می‌شود. نوع تابع انتقال، توسط کاربر تعیین می‌شود که از مشهورترین توابع انتقال می‌توان به نوع خطی، آستانه دو مقداری و سیگموئید اشاره نمود.

علاوه‌براین، ساختار شبکه‌های عصبی مصنوعی می‌تواند تک‌لایه و یا چندلایه باشد.در یک شبکه عصبی تک لایه، الزامی برای یکسان بودن توابع انتقال در هر نورون وجود ندارد یعنی می‌توان از انواع توابع انتقال به‌صورت ترکیبی استفاده نمود.

۳- ارزیابی مدل‌ها

در این تحقیق به‌منظور ارزیابی مدل‌ها شاخص‌های آماری مشخصی پیشنهاد شده است. در این تحقیق از سه شاخص مهم آماری مجذور میانگین مربعات خطا (RMSE)، شاخص ننش-ساتکلیف (NSE)، ضریب هم‌بستگی (R^2) و میانگین درصد قدر مطلق خطا (MAPE) به‌این‌منظور استفاده شده است.

نتایج و بحث

در این تحقیق از روش برنامه‌نویسی ژنتیک برای پیش‌بینی مقدار آب تولیدی روزانه (میزان آب ورودی روزانه به شهر نجف‌آباد ابتدای سال ۱۳۹۴ تا انتهای سال ۱۳۹۸ به تعداد ۱۸۲۶ داده روزانه دبی) استفاده شده و نتایج به‌دست آمده با نتایج مدل شبکه عصبی مصنوعی مقایسه شده است. در این بخش نتایج تحقیق ارائه شده است. در ابتدا به‌منظور انتخاب داده‌های مناسب به‌عنوان ورودی برای مدل‌های پیش‌بینی کننده، از روش ضریب هم‌بستگی پیرسون استفاده شده است، که نتایج آن یک مقدار کمّی بین ۱- تا ۱+ ارائه می‌شود. چنانچه دو بردار از داده‌ها هم‌بستگی در بازه‌های [۱ و ۰/۵] و یا [۰/۵- و ۱-] هم‌بستگی داشته باشند، به‌این معنا است که هم‌بستگی این دو بردار بیش از ۵۰ درصد می‌باشد (مقادیر هم‌بستگی منفی نشان‌دهنده وجود رابطه معکوس بین دو بردار از داده‌ها می‌باشد). مقدار

هم‌بستگی صفر بیانگر این موضوع است که بین دو دسته داده هیچ‌گونه رابطه‌ای وجود ندارد. در این تحقیق به‌منظور بهینه‌سازی حجم داده‌های ورودی و در نتیجه صرفه‌جویی در زمان محاسبات، با تعیین ضریب هم‌بستگی پیرسون، از بردارهای ورودی با هم‌بستگی خارج از محدوده مذکور استفاده نشده است.

در این تحقیق، به‌منظور بررسی اثرات تاخیرهای زمانی بردارهای ورودی بر دقت نتایج مدل پیش‌بینی کننده، اطلاعات مربوط به مقدار آب تولیدی از ۱ تا ۷ روز تاخیر و اطلاعات هواشناسی شامل حداکثر، میانگین و حداقل بارش روزانه (۲۴ ساعته)، حداکثر، میانگین و حداقل رطوبت نسبی روزانه و تبخیر، از ۱ تا ۷ روز تاخیر به‌علاوه داده‌های مربوط به سال گذشته در روز هدف به‌عنوان ورودی منظور و ضریب هم‌بستگی پیرسون آن‌ها محاسبه شد. به‌عبارت‌دیگر، یک سال از داده‌های مشاهداتی در مدل‌سازی استفاده نشده است. بنابراین، ۷ بردار ورودی برای داده‌های مقدار آب تولیدی، ۸ بردار داده ورودی برای داده‌های بارش روزانه، ۸ بردار ورودی برای داده‌های مربوط به حداکثر رطوبت نسبی، ۸ بردار برای داده‌های مربوط به حداقل رطوبت نسبی، ۸ بردار برای داده‌های میانگین رطوبت نسبی و ۸ بردار داده برای میزان تبخیر برای لحاظ کردن تاثیر تاخیر زمانی این پارامترها، بر مقدار آب تولیدی در افق زمانی پیش‌بینی انتخاب شد. سپس، با هدف بررسی میزان تاثیرات بر متغیر خروجی مساله (مقدار آب تولیدی روزانه) آزمون هم‌بستگی پیرسون برای این داده‌ها انجام شد. بررسی نتایج ضریب هم‌بستگی پیرسون بین داده‌های ورودی و بردار خروجی نشان‌دهنده آن است که پارامترهای مقدار ضریب هم‌بستگی پیرسون داده‌های بارش ۲۴ ساعته و تاخیرهای زمانی آن و داده‌های رطوبت با تاخیر زمانی یک ساله، کمتر از ۰/۵ بوده، بنابراین به‌عنوان ورودی مدل‌های این تحقیق منظور نشده و سایر داده‌ها به‌عنوان بردارهای ورودی برای مدل‌های برنامه‌نویسی ژنتیک و شبکه عصبی مصنوعی در نظر گرفته شد.

۱- مدل برنامه‌نویسی ژنتیک (GP)

در حالت کلی، روش GP پارامترهای مختلفی دارد که لازم است مقادیر مطلوب آن‌ها برای هر مدل تعیین شود. مناسب‌ترین مقدار برای هر پارامتر با آنالیز حساسیت و استفاده از روش سعی و خطا تعیین می‌شود. در این تحقیق، ایجاد مدل برنامه‌نویسی ژنتیک در قالب جعبه‌ابزار[۳۷] مربوط به این روش تحت عنوان gplab.toolbox در نرم‌افزار Matlab انجام شده است (Silva و Almeida، ۲۰۰۳). در جدول (۱) پارامترهای مورد استفاده و مقادیر مطلوب آن‌ها در روش برنامه‌نویسی ژنتیک ارائه شده است.

جدول ۱- مقادیر مطلوب پارامترهای روش GP

پارامتر	گزینه‌ها	مقدار مطلوب
انتخاب والدین	Roulette Tournament SUS Lexictour Doubletour (do fitness first) Doubletour (default)	SUS
احتمال وقوع عملگرهای ژنتیک (Crossover, Mutation)	۰/۹۵ ۰/۸ ۰/۷ ۰/۷۵ ۰/۸۵	احتمال آمیزش = ۰/۸ احتمال جهش = ۰/۲
ترتیب‌دهی صف منتخبین	Replace Keepbest Halfelitism Totalelitism	totalelitism
معیار انتخاب از صف منتخبین	Fixedpopsize Resources Pivotfixe	resources
اضافه کردن افراد جدید	M1 M2	M1

در این مدل، داده‌ها به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم شده است. باتوجه‌به ماهیت سری زمانی بودن داده‌ها، ۷۰ درصد داده‌های اول از کل داده‌ها (داده‌های ۱۰۲۲ روز) برای آموزش مدل پیش‌بینی کننده و ۳۰ درصد باقی‌مانده (داده‌های ۴۳۹) به‌عنوان داده‌های آزمایش و صحت‌سنجی استفاده شده است. در این روش، روند مرحله آموزش مشابه الگوریتم تکاملی ژنتیک در تولید نسل جدید می‌باشد. همچنین تمامی داده‌ها نرمال‌سازی شده است. در مدل‌سازی تعداد افراد جمعیت برای هر نسل ۳۰۰ و تعداد نسل‌ها برابر ۱۰۰۰ منظور شده است. با ترکیب حالت‌های مختلف از داده‌های ورودی، مدل‌های متفاوتی پیشنهاد شد که در این تحقیق فقط نتایج بهترین مدل ارائه شده است. باتوجه‌به توضیحات قبلی، در این مدل‌سازی، داده‌های مساله از جمله داده‌های مقدار آب تولیدی روزانه به همراه تاخیرهای ۱ تا ۷ روز و تاخیر یک ساله، داده‌های بیشینه دما به همراه تاخیرهای ۱ تا ۷ روز و تاخیر یک ساله و داده‌های رطوبت نسبی به همراه تاخیرهای ۱ تا ۷روز به‌عنوان ورودی مدل منظور شده است.

در مجموع، باتوجه‌به نتایج آزمون سعی و خطا (آنالیز حساسیت) برای مدل GP، از روش SUS برای انتخاب والد، احتمال آمیزش ۰/۸ و احتمال جهش ۰/۲، روش totalelitism برای ترتیب‌دهی صف منتخبین، معیار resources برای انتخاب از صف منتخبین و روش M1 برای اضافه کردن افراد جدید، استفاده شده است. در ادامه، در جدول (۲) میانگین نتایج شاخص‌های آماری به‌دست آمده از این مدل به‌ازای ۲۰ بار اجرای برنامه با استفاده از روش GP ارائه شده است.

جدول ۲- میانگین نتایج به‌دست آمده با استفاده از مدل GP در ۲۰ بار اجرای مدل

مدل	نوع داده	NSE	RMSE (MCM)	R^2	MAPE
GP	آموزش	۰/۷۲۹	۳۸۱۶/۴۸۴	۰/۷۴۴۵	٪ ۵/۳۸
	آزمایش	۰/۶۳۷۵	۴۴۶۱/۳۰۸	۰/۶۷۷	٪ ۶/۶۷

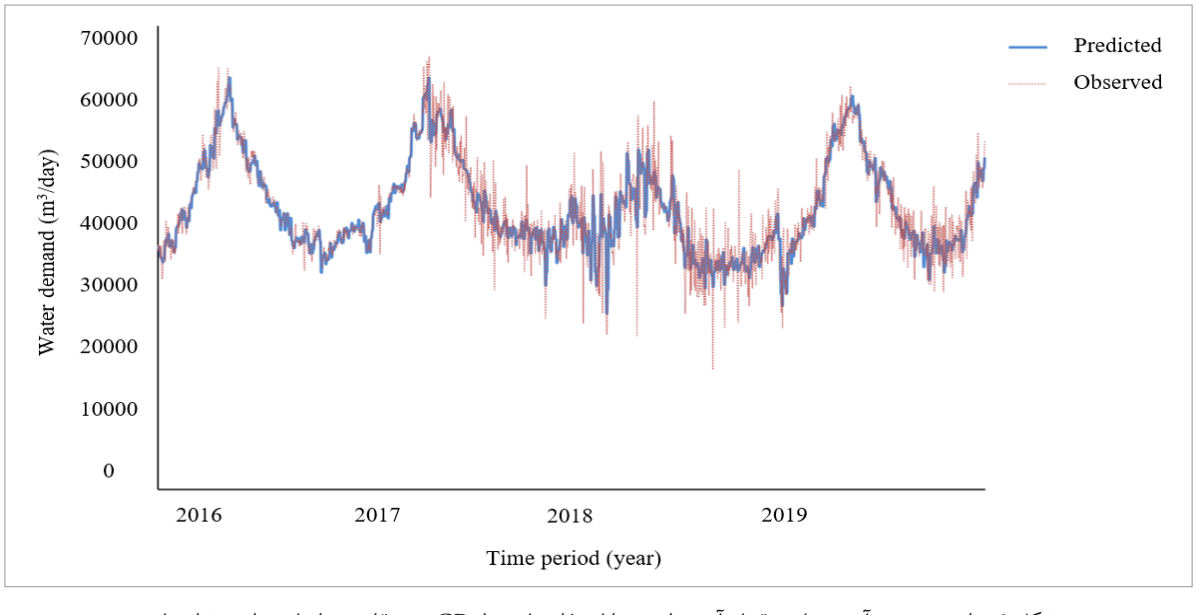
همانطور که توضیح داده شد، روش GP بر خلاف روش شبکه عصبی مصنوعی یک روش جعبه سیاه نبوده و با استفاده از آن، ارائه یک رابطه صریح ریاضی که نشان‌دهنده رابطه بین متغیرهای ورودی و خروجی مساله است برای کاربر امکان‌پذیراست. در این تحقیق با هدف دست‌یابی به دقت مطلوب، پارامترهای روش GP به گونه‌ای تعیین شدند که رابطه ریاضی طولانی و پیچیده حاصل نشود. در این راستا، عملگرهای ریاضی روش GP نیز با انجام آنالیز حساسیت و روش سعی و خطا تعیین شده است. در ادامه، به‌عنوان نمونه یکی از این روابط بین متغیرهای ورودی و خروجی که به‌صورت یک رابطه رگرسیونی میان متغیرهای ورودی (X1, X2, …, X7) با متغیر خروجی (Y) است، در رابطه (۱) ارائه شده است. به‌عبارت‌دیگر، یکی از روابط پیش‌بینی مقدار آب تولیدی بر اساس روش GP مطابق رابطه (۱) است. در این رابطه ریاضی، ورودی‌های X1 تا X7 مقدار آب تولیدی و تاخیرهای ۱ تا ۷ روزه آن و Y میزان آب تولیدی می‌باشند. که این رابطه ریاضی به‌صورت رابطه ساده شده (۲) بازنویسی شده است.

$$Y = plus(minus(minus(plus(X2, X3), mylog(sin(plus(X3, X7)))), mylog(mylog(sin(plus(X5, X1))))),X3) \quad (۱)$$

$$Y = (X2 + X3) - \log(\sin(X3 + X7)) - \log(\log(\sin(X5 + X1))) + X3 \quad (۲)$$

همچنین در شکل (۶)، مقادیر به‌دست‌آمده از مدل GP در مقایسه با داده‌های مشاهداتی ارائه شده است. مقایسه نتایج، نشان‌دهنده عملکرد مناسب روش پیشنهادی در پیش‌بینی میزان مقدار آب تولیدی شهر نجف‌آباد (معیاری از مصرف کل روزانه شهر) می‌باشد. به‌عبارت‌دیگر بررسی نتایج، نشان‌دهنده عملکرد خوب روش به‌ویژه در سال‌های ابتدایی و برای داده‌های آموزش می‌باشد. ولی عملکرد روش برای سال‌های انتهای و به‌ویژه در مقادیر حدی مصرف روزانه آب کاهش یافته است که به دلایل آن در ادامه اشاره شده است.

در مجموع، باتوجه‌به مقادیر به‌دست‌آمده برای شاخص‌های آماری نتایج قابل قبول می‌باشد. در حالت کلی نقطه ضعف روش‌های یادگیری ماشین، عملکرد ضعیف آن در پیش‌بینی مقادیر حدی می‌باشد. علاوه‌براین، مقادیر داده‌های مشاهداتی خطا (نویز) دارد و باید این داده‌ها اصلاح شود. همچنین، در مقادیر حدی آب مصرفی روزانه شهری عوامل دیگری از جمله خرابی و شکستگی لوله و هدررفت آب، تغییرات جمعیت در روزهای خاص و تغییرات دمایی، موثر بوده که بر نتایج تاثیرگذار است.



شکل ۶-نتایج به‌دست‌آمده برای مقدار آب تولیدی با استفاده از مدل GP در مقایسه با داده‌های مشاهداتی

۲- مدل شبکه عصبی مصنوعی

به‌منظور مقایسه عملکرد مدل پیشنهادی برنامه‌نویسی ژنتیک، از مدل شبکه عصبی مصنوعی برای پیش‌بینی مقدار آب تولیدی شهر نجف‌آباد استفاده شد. در حالت کلی شبکه عصبی یک یا چند لایه پنهان و یک لایه خروجی دارد. لایه‌های پنهان در شبکه عصبی توسط نورون‌ها به یکدیگر متصل می‌شوند. تعداد لایه‌های پنهان و تعداد نورون‌های اتصال‌دهنده لایه‌ها در دقت نتایج به‌دستآمده از مدل شبکه عصبی تاثیرگذار است. پارامترهای مذکور، همچنین در مدت زمان انجام محاسبات و رخ‌دادن پدیده بیش‌برازش[۲۸] موثر هستند. بنابراین در این تحقیق با استفاده از روش سعی و خطا تعداد لایه‌های پنهان و تعداد نورون‌های اتصال‌دهنده لایه‌ها تعیین و به‌عبارت‌دیگر آنالیز حساسیت مدل انجام شد. به‌این‌منظور در جدول (۳) مقدار مطلوب پارامترهای مدل شبکه عصبی مصنوعی ارئه شده است.

جدول ۳- مقادیر مطلوب پارامترهای روش شبکه عصبی مصنوعی در ۲۰ باز اجرای مدل

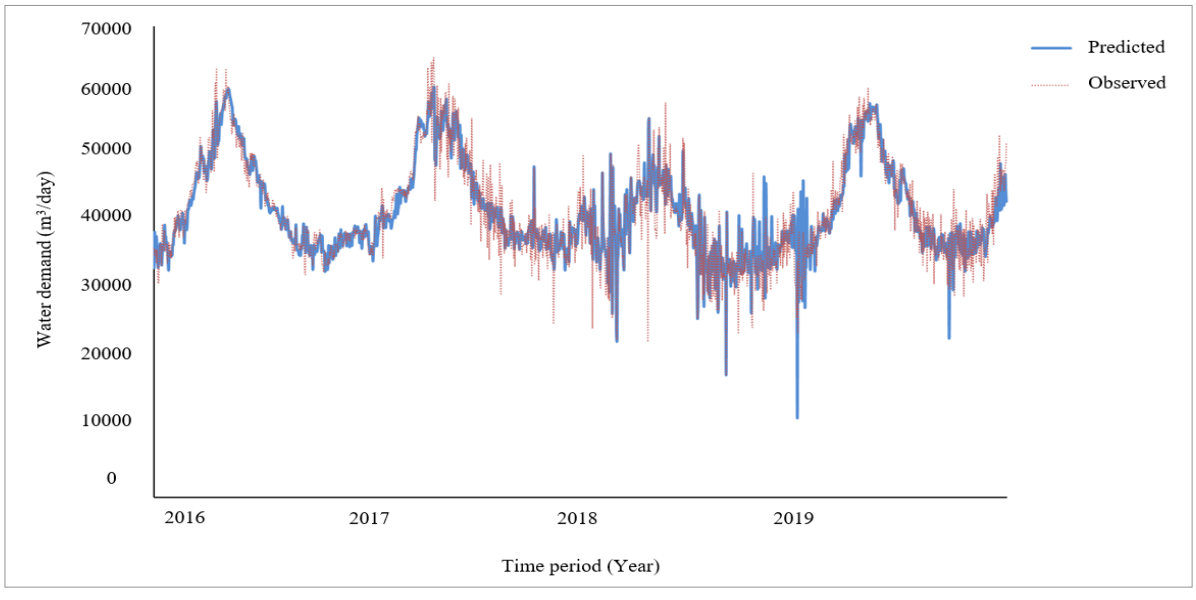
پارامتر	مقدار
تعداد لایه پنهان	۵
تعداد نورون	۳
تابع آموزش	Levenberg –Marquardt
تابع انتقال (لایه پنهان ۱ تا ۵)	Tansig
تابع انتقال (لایه پنهان ۵ به لایه خروجی)	Pureline

در مدل‌سازی با مدل شبکه عصبی مصنوعی باتوجه‌به ماهیت سری زمانی بودن داده‌ها، ۷۰ درصد داده‌های اول از کل داده‌ها (داده‌های ۱۰۲۲ روز) برای آموزش مدل پیش‌بینی کننده و ۳۰ درصد باقی‌مانده (داده‌های ۴۳۹) به‌عنوان داده‌های آزمایش استفاده شده است. با انجام انالیز حساسیت، از الگوریتم Levenberg – Marquardt در مرحله آموزش استفاده شده است. همچنین همه داده‌ها نرمال‌سازی شده است. در این روش نیز مدل‌های مختلف با منظور نمودن ترکیب‌های مختلف از داده‌های ورودی پیشنهاد شده است که در این تحقیق فقط نتایج بهترین مدل که شامل تمامی ۲۴ داده ورودی می‌باشد، ارائه شده است. در جدول (۴) میانگین نتایج به‌دست آمده از ۱۰ مرتبه اجرای مدل شبکه عصبی مصنوعی ارائه شده است. همچنین، مقادیر آب تولیدی پیش‌بینی شده با استفاده از مدل شبکه عصبی مصنوعی در مقایسه با داده‌های مشاهداتی در شکل (۷) نشان داده شده است.

جدول ۴- میانگین نتایج به‌دست آمده با استفاده از روش ANN

مدل	نوع داده	NSE	RMSE (MCM)	R^2	MAPE
ANN	آموزش	۰/۷۶۷	۳۴۶۳/۶۰۹	۰/۷۶۷	٪ ۵/۱۷
	آزمایش	۰/۶۶۱	۴۲۷۰/۸۸۸	۰/۶۸۵	٪ ۶/۹۶

مقایسه نتایج نشان می‌دهد دقت نتایج به‌دست آمده از دو روش برنامه‌نویسی ژنتیک و شبکه عصبی مصنوعی تقریبا مشابه هستند. به‌عبارت‌دیگر، با استفاده از روش شبکه عصبی مصنوعی، دقت بالاتری برای داده‌های آموزش (Train) به‌دست می‌آید ولی با منظور نمودن کل داده‌ها، دقت این روش در مقایسه با روش برنامه‌نویسی ژنتیک کاهش می‌یابد. مهترین دلیل کاهش دقت جواب‌ها این است که مدل شبکه عصبی مصنوعی در پیش‌بینی مقادیر حدی (حداقل و حداکثر) روزانه مصرف آب شهری توانایی ندارد که به دلایل آن پیش از این اشاره شده است. نکته حائز اهمیت که بیانگر برتری روش برنامه‌نویسی ژنتیک بر شبکه عصبی مصنوعی است این است که با استفاده از روش GP ارائه یک رابطه ریاضی صریح برای کاربر امکان‌پذیر است درحالی‌که روش شبکه عصبی مصنوعی یک روش جعبه سیاه[۲۹] بوده و تنها یک بردار از داده‌ها به‌عنوان خروجی در اختیار کاربر قرار می‌گیرد.



شکل ۷- نتایج به دست آمده برای مقدار آب تولیدی با استفاده از مدل ANN با داده‌های مشاهداتی

نتیجه‌گیری

در این تحقیق، از مدل یادگیری ماشین برنامه‌نویسی ژنتیک برای پیش‌بینی مقدار آب تولیدی روزانه (آب مصرفی کل شهری) شهر نجف‌آباد در بازه زمانی سال ۱۳۹۴تا ۱۳۹۸ استفاده شد. در روش پیشنهادی، پارامترهای اندازه‌گیری هواشناسی و داده‌های مقدار آب تولیدی روزانه و تاخیرهای زمانی آنها (۱ تا ۷ روزه و ۳۶۵ روزه) به‌عنوان داده‌های ورودی و میزان کل آب تولیدی در روز هدف به‌عنوان خروجی مدل استفاده شد. با محاسبه ضریب پیرسون داده‌های ورودی موثر تعیین شد. به‌منظور بررسی عملکرد روش برنامه‌نویسی ژنتیک، از مدل شبکه عصبی مصنوعی نیز استفاده شد. مقایسه نتایج نشان داد، روش برنامه‌نویسی ژنتیک توانایی مطلوبی در پیش‌بینی مقدار روزانه آب تولیدی در مقایسه با مدل شبکه عصبی مصنوعی دارد. علاوه‌براین، با استفاده از روش برنامه‌نویسی ژنتیک یک ریاضی صریح بین متغیرهای ورودی و خروجی مساله به‌منظور پیش‌بینی مقادیر روزانه آب مصرفی شهری ارائه شد. این رابطه به مدیران و تصمیم‌گیران حوزه مدیریت شبکه توزیع آب شهری کمک خواهد نمود. مهمترین محدودیت‌های تحقیق حاضر عدم دسترسی به داده‌های مورد نیاز برای برآورد نیاز آبی شهر ازجمله داده‌های مصرف قرائت شده از کنتور مشترکین، هدررفت آب و جمعیت مشترکین در نواحی مختلف شهر و محدودیت مدل‌ها در پیش‌بینی مقادیر حدی مصرف روزانه بود. در ادامه تحقیق حاضر، استفاده از سایر مدل‌های هوش مصنوعی و یادگیری ماشین در پیش‌بینی مصرف و نیاز آبی، تکمیل داده‌ها و تغییر بازه زمانی پیش‌بینی مصرف به ماهانه و فصلی پیشنهاد می‌شود.

پی‌نوشت‌ها

1-Meta Regression
2-Support Vector Regression
3-Fourier Time Series
4-Franca
5-Fixed Regression Stricture
6-Fixed Regression Stricture
7-Pattern-based model
8-Moving window technique
9-Probabilistic Markov chain-based model
10-Naïve
11-Padua
12-Exter
13-Gradient boosting machines
14-AutoRegressive Integrated Moving Average
15-Support Vector Regression
16-Random Forest
17-Extremely randomized trees
18-Beijing–Tianjin–Hebei Region
19-Gradient boosting decision trees
20-Long short-term memory networks

21-Artificial neural network
22-Genetic algorithm
23-Syntax tree
24-Terminal
25-Function
26-Neuron
27-Toolbox
28-Overfitting
29-Black box

منابع

Barati, R., Salehi Neyshabouri, S. A. A., & Ahmadi, G. (2014). Development of empirical models with high accuracy for estimation of drag coefficient of flow around a smooth sphere: An evolutionary approach. Powder Technology, 257, 11-19. doi: 10.1016/j.powtec.2014.02.045

Brentan, B. M., Luvizotto Jr, E., Herrera, M., Izquierdo, J., & Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. Journal of Computational and Applied Mathematics, 309, 532-541. doi: 10.1016/j.cam.2016.02.009

Goodarzi, M., Abedi-Koupai, J., Heidarpour, M., & Safavi, H. R. (2016). Evaluation of the Effects of Climate Change on Groundwater Recharge Using a Hybrid Method. Water Resources Management, 30(1), 133-148. doi: 10.1007/s11269-015-1150-4

Kazemi, M., & Barati, R. (2022). Application of dimensional analysis and multi-gene genetic programming to predict the performance of tunnel boring machines. Applied Soft Computing, 124, 108997. doi: 10.1016/j.asoc.2022.108997

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of LSTM Networks for Water Demand Prediction in Optimal Pump Control. Water, 13(5), 644. doi: 10.3390/w13050644

Lee, M., Tansel, B., & Balbin, M. (2011). Influence of residential water use efficiency measures on household water demand: A four year longitudinal study. Resources, Conservation and Recycling, 56(1), 1-6. doi: 10.1016/j.resconrec.2011.08.006

Pacchin, E., Gagliardi, F., Alvisi, S., & Franchini, M. (2019). A Comparison of Short-Term Water Demand Forecasting Models. Water Resources Management, 33(4), 1481-1497. doi: 10.1007/s11269-019-02213-y

Polebitski Austin, S., & Palmer Richard, N. (2010). Seasonal Residential Water Demand Forecasting for Census Tracts. Journal of Water Resources Planning and Management, 136(1), 27-36. doi: 10.1061/(ASCE)WR.1943-5452.0000003

Romano, G., Salvati, N., & Guerrini, A. (2014). Estimating the Determinants of Residential Water Demand in Italy. Water, 6, 2929–2945. doi: 10.3390/w6102929

Sebri, M. (2016). Forecasting urban water demand: A meta-regression analysis. Journal of Environmental Management, 183, 777-785. doi: 10.1016/j.jenvman.2016.09.032

Shabani, S., Candelieri, A., Archetti, F., & Naser, G. (2018). Gene expression programming coupled with unsupervised learning: a two-stage learning process in multi-scale, short-term water demand forecasts. Water, 10(2), 142. doi: 10.3390/w10020142

Shuang, Q., & Zhao, R. T. (2021). Water Demand Prediction Using Machine Learning Methods: A Case Study of the Beijing–Tianjin–Hebei Region in China. Water, 13(3), 310. doi: 10.3390/w13030310

Silva, S., & Almeida, J. (2003). Gplab–a genetic programming toolbox for matlab. In: Proceedings of the Nordic MATLAB Conference. Copenhagen, Denmark.

Smolak, K., Kasieczka, B., Fialkiewicz, W., Rohm, W., Siła-Nowicka, K., & Kopańczyk, K. (2020). Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water Journal, 17(1), 32-42. doi: 10.1080/1573062X.2020.1734947

Suero Francisco, J., Mayer Peter, W., & Rosenberg David, E. (2012). Estimating and Verifying United States Households' Potential to Conserve Water.

Journal of Water Resources Planning and Management, 138(3), 299-306. doi: 10.1061/(ASCE)WR.1943-5452.0000182

Xenochristou, M., & Kapelan, Z. (2020). An ensemble stacked model with bias correction for improved water demand forecasting. Urban Water Journal, 17(3), 212-223. doi: 10.1080/1573062X.2020.1758164