



Tabriz Daily Rainfalls Modeling via Hybridized Tree Based and Seasonal-Trend Component Bagging Method

S. Javidan¹, M.T. Sattari ^{2*}, Sh. Mohsenzadeh³

Received: 16-05-2022

Revised: 08-06-2022

Accepted: 06-07-2022

Available Online: 22-09-2022

How to cite this article:

Javidan S., Sattari M.T., and Mohsenzadeh Sh. 2022. Tabriz Daily Rainfalls Modeling via Hybridized Tree Based and Seasonal-Trend component Bagging Method. Journal of Water and Soil 36(3): 407-420. (In Persian with English abstract)

DOI: [10.22067/JSW.2022.76512.1161](https://doi.org/10.22067/JSW.2022.76512.1161)

Introduction

Precipitation is one of the most important components of water cycle. Accurate precipitation measurement is essential for flood forecasting and control, drought analysis, runoff modeling, sediment control and management, watershed management, agricultural irrigation planning, and water quality studies. Determining the correct amount of precipitation in cities and rural areas is also important for managing floods. The precipitation process is completely non-linear and involves randomness in terms of time and space. Therefore, it is not easy to explain that with simple linear models due to various climatic factors and may contain major errors. Therefore, various methods and models have been proposed to evaluate, and predict precipitation. This study aimed to estimate the daily precipitation of Tabriz based on hybridized tree-based and Bagging methods by using neighboring stations.

Materials and Methods

In the present study, the rainfall data of adjacent stations in Urmia lake basin (Sahand, Sarab, Urmia, Maragheh and Mahabad) were employed in 1986-2021 to estimate the daily rainfall in Tabriz. About 70% of data were considered for calibration and 30% of data were applied for validation. Using the correlation matrix and Relief algorithm, various input components were identified. Modeling was performed using tree-based data mining methods including M5P, RT and REPT and Bagging method. The daily precipitations of Tabriz was decomposed into their components by seasonal-trend analysis method. Its components, including trend, seasonal and residual, were used in different input scenarios to investigate the effect of these components on improving the modeling results. To evaluate the modeling performance, the indices of correlation coefficient, Root Mean Square Error, Nash-Sutcliffe Efficiency and modified Wilmut coefficient were applied.

Results and Discussion

RT and REPT methods increased the accuracy of the model and decreased its error when they were used as the basic algorithm of the Bagging method. This was not the case with the M5P method, as the results were slightly weaker. It was also observed that Tabriz rainfall is largely influenced by Sahand rainfall, as the most models gave reliable estimates by using the rainfall data for Sahand station. This can be explained by the high correlation between Tabriz rainfall and Sahand. The results showed that the first scenario (Sahand) for M5P, RT, REPT and B-M5P method, the fifth scenario (Sahand, Sarab, Urmia, Maragheh and Mahabad) for the B-RT method, and the fourth scenario (Sahand, Sarab, Urmia and Mahabad) for the B-REPT method were the best scenarios. The best performance was found for the scenario 1 of the M5P decision tree model, followed by the Bagging method with the M5P base algorithm. In general, it was concluded that application of the Bagging method produced reliable results. Modeling without considering the decomposition components was compared with modeling with decomposition components. Adding seasonal, trend and residual components to the modeling input combinations significantly improved the accuracy of the results. Application of Bagging method in most cases also increased

1, 2 and 3- M.Sc. Student, Associate Professor and M.Sc. Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, respectively.

(*- Corresponding Author Email: mtsattar@tabrizu.ac.ir)

the modeling accuracy. The first scenario (Sahand and residual) for M5P and B-M5P methods, the tenth scenario (residual, trend, seasonal, Sahand and Sarab) for RT, REPT and B-REPT methods, and the eighth scenario (residual, trend and Sahand) for B-RT method were selected as the best scenarios. As a result, among the stations, Sahand, due to proximity and high correlation, and Sarab, due to greater correlation, had a great impact on precipitation in Tabriz. In general, the Bagging method with the basic M5P algorithm (B-M5P) was best suited in the first scenario. Thus, adding precipitation analysis components and using the Bagging method improve the modeling results with tree-based data mining methods.

Conclusion

Our results showed that Bagging method provided acceptable results in most cases. In the first case, the first scenario of M5P method including Sahand precipitation data was selected as the superior method and scenario. As a result, Sahand was the most effective station in estimating Tabriz rainfall with the highest correlation and the shortest distance from Tabriz. In the second case, with the decomposition components, the accuracy of the results increased significantly. The Bagging method with the basic M5P algorithm, the parameters of Sahand precipitation and the residual of Tabriz precipitation was considered as the best modeling algorithm. It can be concluded that using Bagging method and decomposition components with the closest station to the studied station results in the highest accuracy. Therefore, Bagging models with tree-based algorithm can be considered as simple and widely used methods.

Keywords: Bagging method, Decomposition, Modified Wilmot, Tree models, Urmia lake basin

مقاله پژوهشی

جلد ۳۶، شماره ۳، مرداد-شهریور ۱۴۰۱، ص. ۴۲۰-۴۰۷

مدل سازی بارش روزانه تبریز با روش های درختی ادغام شده با تجزیه فصلی-روند و رویکرد دسته بندی

سحر جاویدان^۱ - محمدتقی ستاری^{۲*} - شکوه محسن زاده^۳

تاریخ دریافت: ۱۴۰۱/۰۲/۲۶

تاریخ پذیرش: ۱۴۰۱/۰۴/۱۵

چکیده

بارش به عنوان یک متغیر تصادفی با داشتن تغییرات مکانی و زمانی یکی از عناصر پیچیده در چرخه هیدرولوژی است. هدف پژوهش حاضر برآورد میزان بارش روزانه تبریز در بازه زمانی ۳۶ ساله (۱۹۸۶-۲۰۲۱) با استفاده از گروه روش های درختی شامل، مدل درختی M5P، درخت تصادفی، کاهش خطای هرس درخت و روش دسته بندی است. بدین منظور از مقادیر بارش ایستگاه های حوضه دریاچه ارومیه از جمله سهند، سراب، ارومیه، مراغه و مهاباد در ترکیب های ورودی مختلف استفاده شد. ماتریس همبستگی و الگوریتم رلیف مبنای انتخاب سناریوهای ورودی در نظر گرفته شد و تأثیر مؤلفه های تجزیه فصلی-روند در بهبود نتایج مدل سازی بررسی شد. عملکرد روش های مذکور با معیارهای ضریب همبستگی، ریشه میانگین مربعات خطا، ضریب نش ساتکلیف، میانگین خطای قدر مطلق و ضریب ویلموت اصلاح شده مورد ارزیابی قرار گرفت. بررسی نتایج نشان داد رویکرد دسته بندی در اکثر موارد نتایج قابل قبولی ارائه نموده و باعث بهبود نتایج مدل سازی می گردد. بررسی ها مشخص نمود که ایستگاه سهند با بیشترین همبستگی و کمترین فاصله از تبریز، مؤثرترین ایستگاه مجاور در برآورد میزان بارش تبریز می باشد. در حالت اول و بدون اعمال مؤلفه های تجزیه (روند، فصلی و باقیمانده) در بین روش های مورد استفاده روش M5P با سناریو اول شامل بارش سهند به عنوان روش و سناریو برتر انتخاب شد. در حالت دوم با وارد شدن مؤلفه های تجزیه، دقت تخمین ها به صورت چشم گیری افزایش یافت. ادغام روش دسته بندی با الگوریتم پایه M5P با پارامترهای بارش سهند و باقیمانده بارش تبریز با $NS=0.95$ و $R=0.98$ به عنوان برترین حالت انتخاب گردید. در حالت کلی نتایج نشان داد، بهره گیری توأم از رویکرد دسته بندی مدل ها و الگوریتم پیش پردازش مؤلفه های تجزیه باعث بهبود نتایج مدل سازی بارش روزانه تبریز می شود. به طوریکه مقدار خطای RMSE نسبت به حالت اول ۶۰/۶۴ درصد کاهش یافت. بنابراین به علت استفاده از حداقل تعداد پارامتر ورودی و ارائه نتایج قابل قبول، مدل های دسته بندی با الگوریتم پایه درختی به عنوان روش های ساده و پر کاربرد پیشنهاد می گردد.

واژه های کلیدی: تجزیه، حوضه دریاچه ارومیه، رویکرد دسته بندی، مدل های درختی، ویلموت اصلاح شده

مقدمه

کشاورزی و مطالعات کیفیت آب مهم و ضروری است. تشخیص دقیق بارندگی در شهرها و مناطق روستایی برای پیشگیری و مدیریت مخاطرات ناشی از سیل اهمیت دارد (Sattari et al., 2020). با توجه به این که فرآیند بارش، فرآیندی کاملاً غیرخطی و از نظر زمانی و مکانی تصادفی است و عوامل مختلف اقلیمی مانند شکل، شدت،

بارندگی یکی از مهم ترین ورودی های سیستم های آبی است و نقش کلیدی در تمام مطالعات و مدیریت منابع آب دارد. اندازه گیری دقیق بارش در پیش بینی و کنترل سیل، تحلیل خشکسالی، مدل سازی جریان، کنترل و مدیریت رسوب، مدیریت حوضه، برنامه ریزی آبیاری

۱، ۲ و ۳- به ترتیب دانشجوی کارشناسی ارشد، دانشیار و دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

(Email: mtsattar@tabrizu.ac.ir)

(*- نویسنده مسئول)

چشمگیر کارایی مدل درخت تصمیم می‌شود. نگاهامولا و همکاران (Nagahamulla et al., 2014) دو مدل مجموعه شبکه عصبی با خوشه‌بندی K (ENN-K^۲) و مجموعه شبکه عصبی با الگوریتم ژنتیک (ENN-GA^۳) را با روش‌های دسته‌بندی و تقویتی مقایسه کردند. آن‌ها دریافتند که مدل ENN-GA با خطای نسبتاً کمتری دقیق‌تر از مدل‌های Bagging و Boosting است. بوشارا و ابراهیم (Bushara and Abraham, 2015) از روش‌های ترکیبی برای پیش‌بینی بلندمدت بارندگی استفاده کردند. نتایج به‌دست آمده از تحقیق آن‌ها نشان داد که روش ترکیبی بهترین عملکرد را در مقایسه با الگوریتم‌های پایه دارد. روش ترکیبی نه تنها دقت پیش‌بینی را افزایش می‌دهد، بلکه منجر به اطمینان بیشتر در نتایج می‌شود. میسرا و همکاران (Mishra et al., 2017) مطالعه جامعی از تکنیک‌های آماری و روش‌های داده‌کاوی برای تجزیه و تحلیل داده‌های سری زمانی ارائه دادند. آن‌ها نتیجه گرفتند که تکنیک‌های آماری و تکنیک‌های مبتنی بر شبکه عصبی را می‌توان به صورت موازی در تجزیه و تحلیل داده‌های سری زمانی استفاده کرد و با کمک تکنیک‌های محاسباتی نرم، نتایج بهتری ممکن است یافت شود. چوبین و همکاران (Choubin et al., 2018) به پیش‌بینی بارش با استفاده از مدل طبقه‌بندی و رگرسیون درختان پرداختند. آن‌ها نتیجه گرفتند که CART^۴ با ضریب نش-ساتکلیف بیشتر از ۰/۷۵ نتایج بهتری در مقایسه با ANFIS^۵ و ARIMA^۶ در پیش‌بینی بارش دارد. ادنان و همکاران (Adnan et al., 2021) به مقایسه روش‌های دسته‌بندی^۷ و تقویت^۸ برای طبقه‌بندی بارش در ایستگاه هواشناسی سلطان سیاریف قاسم دوم در پکان بارو پرداختند. آن‌ها دریافتند که تقویت گرادیان تصادفی^۹ با مدل‌سازی تقویت گرادیان بهترین الگوریتم بر اساس دقت متوسط است. همچنین این الگوریتم در مقایسه با سایر الگوریتم‌ها کاربرد بیشتری دارد. یو و هاسکینز (Yu and Haskins, 2021) برای پیش‌بینی بارش منطقه‌ای در ایالت نیویورک از روش‌های یادگیری ماشینی استفاده کردند. در تحقیق آن‌ها مدل KNN^{۱۰} هم برای طبقه‌بندی و هم برای رگرسیون به‌عنوان مدلی با دقت بالا و خطای کمتر معرفی شد. لیو و ملیز (Liyew and Melese, 2021) روش‌های رگرسیون خطی چندمتغیره، جنگل تصادفی و افزایش گرادیان شدید^{۱۱} را برای پیش‌بینی میزان بارندگی روزانه به کار بردند. آن‌ها دریافتند که الگوریتم یادگیری ماشینی

مدت، دوره بارش، تبخیر و غیره در آن نقش دارند، تشریح آن با مدل‌های ساده و خطی به راحتی امکان‌پذیر نبوده و با خطای زیادی همراه است. بنابراین، روش‌ها و مدل‌های گوناگونی جهت ارزیابی، پیش‌بینی و تخمین بارندگی ارائه شده است (Omidvar and Azhdarpoor, 2013). روش‌های مبتنی بر داده، قابلیت در نظر گرفتن رفتار غیرخطی متغیرهای هواشناسی و هیدرولوژیکی حوضه‌ها را بدون توجه کافی بر شرایط فیزیکی حاکم بر مسئله دارا هستند. در سال‌های اخیر مدل‌های درختی به‌عنوان یکی از تکنیک‌های داده‌کاوی در پیش‌بینی پارامترهای هیدرولوژیکی و هواشناسی مورد توجه قرار گرفته‌اند. یافتن یک الگوریتم یادگیری ماشینی با کارایی بالا ممکن است کمی طول بکشد. این به این دلیل است که یادگیری ماشین کاربرد، پر از فرآیند آزمون و خطا است. رویکرد ترکیبی^۱ که چندین مدل مختلف را ترکیب می‌کند، یکی از راه‌های بهبود دقت نتایج مدل‌سازی است (Yobero, 2018). برخلاف سایر روش‌های طبقه‌بندی که فقط از یک الگوریتم طبقه‌بندی استفاده می‌کنند، روش ترکیبی مجموعه‌ای از الگوریتم‌ها را ادغام و نتایج را برای تصمیم‌گیری ترکیب می‌کند (Zhou, 2012). بررسی منابع نشان می‌دهد تحقیقات زیادی در زمینه برآورد میزان بارش با استفاده از روش‌های داده‌منا در سطح ملی و بین‌المللی انجام یافته است. دستورانی و همکاران (Dastourani et al., 2013) کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه سینوپتیک یزد را بررسی کردند. نتایج تحقیق آن‌ها نشان داد که در ایستگاه یزد، مدل درخت تصمیم‌گیری خصوصاً در شرایطی که از میانگین متحرک ۵ ساله داده‌ها استفاده گردد، دارای توانایی مناسبی در پیش‌بینی میزان بارش می‌باشد. سناری و نهرین (Sattari and Nahrein, 2014) مقادیر حداکثر بارش روزانه ایستگاه‌های اهر و جلفا را با استفاده از شبکه‌های عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 پیش‌بینی کردند. نتایج تحقیق آن‌ها نشان داد که مدل درختی M5 به دلیل ارائه روابط خطی ساده و قابل فهم، می‌تواند به‌عنوان روشی کاربردی و جایگزین برای محاسبه حداکثر بارش روزانه در ماه، مورد توجه قرار گیرد. امیدوار و همکاران (Omidvar et al., 2014) کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه سینوپتیک کرمانشاه را ارزیابی نمودند. آن‌ها نتیجه گرفتند که درخت تصمیم‌گیری رگرسیونی، مدلی نسبتاً کارا در پیش‌بینی بارش است و استفاده از میانگین متحرک منجر به افزایش

- 8- Boosting
- 9- Stochastic Gradient Boosting
- 10- K-Nearest Neighbors
- 11- Extreme Gradient Boosting

- 1- Ensemble
- 2- Ensemble Neural Network with K-means
- 3- Ensemble Neural Network with Genetic Algorithms
- 4- Classification And Regression Trees
- 5- Adaptive Neuro-Fuzzy Inference System
- 6- Auto Regressive Integrated Moving Average
- 7- Bagging

به دامنه کوه سهند و از غرب به جلگه تبریز و دریاچه ارومیه محدود است (Asakereh and Akbarzadeh, 2017). موقعیت جغرافیایی ایستگاه‌های سینوپتیک و ویژگی‌های آماری بارش در جدول ۱ و موقعیت مکانی ایستگاه تبریز و فاصله آن از ۵ ایستگاه مجاور و همبستگی بارش بین آن‌ها در شکل ۱ نشان داده شده است.

در پژوهش حاضر برای برآورد میزان بارش روزانه تبریز از مقادیر بارش ایستگاه‌های مجاور (سهند، سراب، ارومیه، مراغه و مهاباد) در دوره آماری ۳۶ ساله (۱۹۸۶-۲۰۲۱) استفاده شد. از بین داده‌های موجود، ۷۰٪ برای واسنجی و ۳۰٪ برای صحت‌سنجی در نظر گرفته شد. با استفاده از ماتریس همبستگی و الگوریتم رلیف انواع ترکیب‌های ورودی مدل‌ها در قالب چندین سناریو مشخص شد (جدول ۲). رلیف یک الگوریتم انتخاب ویژگی برای کاهش ابعاد مسئله است که اولین بار توسط کیرا و رندل (Kira and Rendell, 1992) پیشنهاد گردید. نکات قوت این الگوریتم ساده بودن اصول و عدم پیچیدگی آن، قابل حل بودن با توابع چندجمله‌ای مرتبه پایین، قابل استفاده بودن برای داده‌های پیوسته و نیاز به تعداد کم داده‌های آموزشی است. در یک مجموعه داده با تعداد N نمونه (داده مشاهداتی) و تعداد P ویژگی که مربوط به دو طبقه مختلف هستند، هر ویژگی باید در بازه $(0, 1)$ قرار گیرد. الگوریتم مذکور m بار تکرار شده و در هر مرتبه از یک بردار وزنی متفاوت که از صفر شروع می‌گردد، استفاده می‌کند. در هر تکرار، الگوریتم مذکور بردار ویژگی X را که متعلق به یک نمونه تصادفی است و بردارهای ویژگی نزدیکترین نمونه به نمونه X در طبقه مورد نظر را توسط تابع فاصله اقلیدسی^۷ انتخاب می‌کند. پس از m تکرار هر یک از عناصر بردار وزن توسط m تقسیم‌بندی می‌شوند. نتیجه این عمل به دست آمدن یک بردار مرتبط است که اگر مقدار بردار مرتبط یک ویژگی از آستانه تعریف شده بیشتر گردد، آن ویژگی انتخاب می‌گردد.

همچنین برای تجزیه و تحلیل داده‌ها، نمودار طرح خطی برای تمام ایستگاه‌ها در شکل ۲ آورده شد که از طریق روش طرح‌ریزی خطی با تجزیه و تحلیل داده‌های اکتشافی و با استفاده از نرم‌افزار Orange تهیه شده است. هر نقطه روی نمودار محل تقاطع بارش تمام ایستگاه‌ها است. همان‌طور که از شکل ۲ مشاهده می‌شود، به‌جز برخی از مقادیر شدید باران که به طور طبیعی ظاهر می‌شوند، سایر داده‌های بارش در یک خوشه برای هر یک از ایستگاه‌ها قرار می‌گیرند.

تقویت گرادیان شدید بهتر از سایر الگوریتم‌ها عمل می‌کند. بررا و همکاران (Barrera et al., 2022) به تحلیل مقایسه‌ای الگوریتم‌های یادگیری ماشین مدرن برای پیش‌بینی سری‌های زمانی بارندگی پرداختند. آن‌ها نتیجه گرفتند در بین تمام مدل‌های آزمایش شده، شبکه حافظه بلند مدت کوتاه انباشته^۱ با دو لایه پنهان و شبکه حافظه بلند مدت کوتاه دو جهته^۲ بهترین عملکرد را داشتند. کابزولو (Cabezuelo, 2022) برای پیش‌بینی بارندگی در استرالیا از روش‌های k نزدیکترین همسایگی، درخت تصمیم، جنگل تصادفی و شبکه عصبی استفاده کردند. نتایج تحقیق آن‌ها نشان داد که شبکه عصبی بالاترین دقت را در بین روش‌های مورد مطالعه داشت. اندالی و همکاران (Endalie et al., 2022) برای پیش‌بینی بارش روزانه جیما، منطقه‌ای واقع در جنوب غربی ارومیا روش‌های پرسپترون چندلایه^۳، k نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، درخت تصمیم و روش یادگیری عمیق LSTM^۴ را به کار بردند. نتایج آن‌ها نشان داد که مدل LSTM دارای RMSE کمتر و R^2 بالاتر است. هدف پژوهش حاضر برآورد میزان بارش روزانه تبریز با استفاده از مقادیر بارش ایستگاه‌های مجاور در دوره آماری ۳۶ سال (۱۹۸۶-۲۰۲۱) است. بدین منظور از روش‌های داده‌کاوی مبتنی بر درخت شامل درخت تصادفی (RT^5)، مدل درختی MSP و کاهش خطای هرس درخت (REPT^۶) استفاده گردید. ارزیابی دقت و عملکرد رویکرد ادغام دسته‌بندی به‌عنوان روشی ترکیبی و الگوریتم پیش پردازش تجزیه فصلی-روند در راستای بهبود نتایج مدل‌سازی یکی دیگر از اهداف این مطالعه می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه و داده‌های مورد استفاده

حوضه آبریز دریاچه ارومیه بین مختصات ۴۴ درجه و ۱۴ دقیقه تا ۴۷ درجه و ۵۳ دقیقه طول شرقی و ۳۵ درجه و ۴۰ دقیقه تا ۳۸ درجه و ۳۰ دقیقه عرض شمالی واقع شده است. میانگین بارش روزانه تبریز در ۳۶ سال اخیر ۰/۷۱۱ میلی‌متر و حداقل و حداکثر مقدار بارش به ترتیب ۰ و ۵۸ میلی‌متر بوده است. در این حوضه از مناطق مرکزی حوضه به سوی مناطق مرتفع پیرامونی بر مقدار بارندگی افزوده می‌شود (Tahroudi et al., 2017). هدف پژوهش حاضر تخمین میزان بارش روزانه تبریز در دوره آماری ۳۶ ساله (۱۹۸۶-۲۰۲۱) می‌باشد. تبریز مرکز استان آذربایجان شرقی بوده و دارای مساحتی حدود $1200 \text{ (km}^2\text{)}$ می‌باشد. تبریز از شمال به کوه عینالی، از جنوب

5- Random Tree

6- Reduced Error Pruning Tree

7- Euclidean Distance Function

1- Stacked- Long-Short Term Memory

2- Bidirectional- Long-Short Term Memory

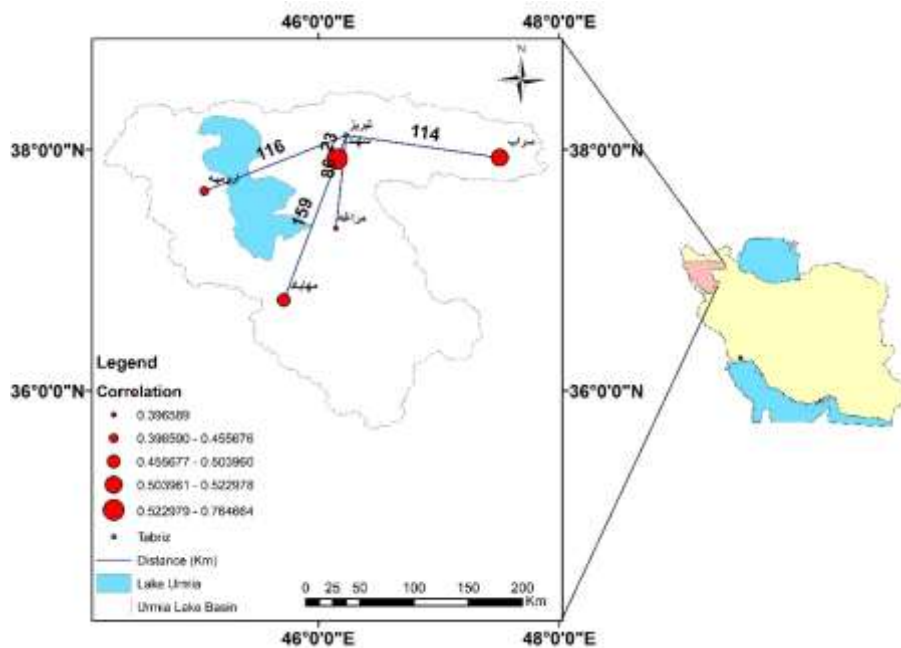
3- Multilayer Perceptron

4- Long Short Term Memory

جدول ۱- مشخصات آماری داده‌های بارش و موقعیت جغرافیایی ایستگاه‌های منتخب در دوره آماری ۳۶ ساله (۱۹۸۶-۲۰۲۱)

Table 1- Statistical characteristics of precipitation data and geographical location of selected stations in the statistical period of 36 years (1986-2021)

ایستگاه (Station)	حداقل (میلی متر) Minimum (mm)	حداکثر (میلی متر) Maximum (mm)	میانگین (میلی متر) Mean (mm)	انحراف معیار (میلی متر) Standard Deviation (mm)	طول جغرافیایی (Longitude)	عرض جغرافیایی (Latitude)	ارتفاع (متر) Elevation (m)
تبریز (Tabriz)	0	58	0.71	2.50	46° 14' 1"	38° 7' 20"	1345
سهند (Sahand)	0	49	0.57	2.22	46° 9' 24"	37° 55' 25"	1695
سراب (Sarab)	0	35	0.64	2.21	47° 30' 28"	37° 56' 6"	1680
ارومیه (Urmia)	0	64	1.05	4.36	45° 3' 19"	37° 39' 29"	1328
مرآغه (Maraghe)	0	79	1.24	4.33	46° 8' 46"	37° 20' 51"	1342
مهاباد (Mahabad)	0	68	1.07	3.88	45° 42' 55"	36° 45' 11"	1352



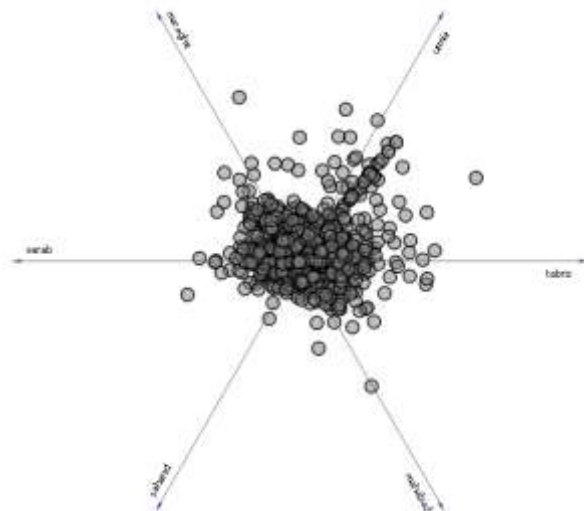
شکل ۱- موقعیت مکانی ایستگاه‌های مورد مطالعه به همراه همبستگی و فاصله بین ایستگاه هدف با سایر ایستگاه‌ها

Figure 1- Location of the studied stations along with the correlation and distance between the target and other stations

جدول ۲- روش‌های انتخاب سناریو و پارامترهای دخیل در آن بدون در نظر گرفتن الگوریتم مؤلفه‌های تجزیه فصلی-روند

Table 2- Scenario Selection Methods and Involved Parameters without Seasonal-Process Analysis Components

شماره سناریو (Scenario Number)	ورودی (Inputs)	روش انتخاب سناریو (Scenario selection method)
1	Sahand	Correlation Matrix
2	Sahand, Sarab	Correlation Matrix
3	Sahand, Sarab, Mahabad	Correlation Matrix
4	Sahand, Sarab, Mahabad, Urmia	Correlation Matrix
5	Sahand, Sarab, Mahabad, Urmia, Maraghe	Correlation Matrix
6	Sahand, Sarab, Urmia	Relief



شکل ۲- نمودار طرح خطی بارش ایستگاه‌های مورد مطالعه
Figure 2- Rainfall Linear Projection Graph of the studied stations

جدول ۳- روش‌های انتخاب سناریو و پارامترهای دخیل در آن با در نظر گرفتن مؤلفه‌های تجزیه فصلی-روند

Table 3- Scenario selection methods and parameters involved in it considering the components of seasonal-trend analysis

شماره سناریو (Scenario Number)	ورودی (Inputs)	روش انتخاب سناریو (Scenario selection method)
1	Resid, Sahand	Correlation Matrix
2	Resid, Sahand, Sarab	
3	Resid, Sahand, Sarab, Mahabad	
4	Resid, Sahand, Sarab, Mahabad, Urmia	
5	Resid, Sahand, Sarab, Mahabad, Urmia, Maraghe	
6	Resid, Sahand, Sarab, Mahabad, Urmia, Maraghe, Seasonal	
7	Resid, Sahand, Sarab, Mahabad, Urmia, Maraghe, Seasonal, Trend	
8	Resid, Trend, Sahand	Relief
9	Resid, Trend, Sahand, Sarab	
10	Resid, Trend, Sahand, Sarab, Seasonal	
11	Resid, Trend, Sahand, Sarab, Seasonal, Urmia	
12	Resid, Trend, Sahand, Sarab, Seasonal, Urmia, Mahabad	

باقیمانده، نویز یا تصادفی بودن را نشان می‌دهد. به‌عنوان مثال در سناریو اول که بر اساس همبستگی انتخاب شده است برای پیش‌بینی بارش روزانه ایستگاه سینوپتیک تبریز از مقادیر بارش روزانه ایستگاه سهند و یکی از مؤلفه‌های تجزیه بارش تبریز (باقیمانده) استفاده شد.

روش‌های مورد مطالعه

درخت تصادفی

درخت تصادفی یک طبقه‌بندی‌کننده تحت نظارت است و از یک ایده جمع‌آوری برای تولید مجموعه‌ای تصادفی از داده‌ها برای ساخت درخت تصمیم استفاده می‌کند. درختان تصادفی اساساً ترکیبی از دو

مدل‌سازی با استفاده از گروه مدل‌های درختی شامل M5P، RT و REPT و روش دسته‌بندی انجام شد. بارش‌های روزانه تبریز با روش تجزیه فصلی-روند به سری اجزای خود تجزیه شدند و از مؤلفه‌های آن که شامل روند^۱، فصلی^۲ و باقیمانده^۳ بودند در ترکیب‌های ورودی مختلف استفاده شد (جدول ۳) تا تأثیر وجود این مؤلفه‌ها در بهبود نتایج مدل‌سازی بررسی شود. روند بارش تبریز با استفاده از نرم‌افزار Xlstat و آزمون من-کندال^۴ مورد تست قرار گرفت و با قبول فرض صفر عدم وجود روند در سطح معنی‌داری ۵ درصد مشخص گردید. تجزیه فصلی-روند هم در نرم‌افزار پایتون انجام گرفت. روند، دنباله‌ای از نقاط داده را نشان می‌دهد که در یک محور زمانی پیوسته متفاوت است، در حالی که فصلی بودن یک الگوی پیوسته و منظم است و در یک بازه زمانی ثابت تکرار می‌شود.

3- Residual
4- Mann- Kendall Tests

1- Trend
2- Seasonal

روش‌های یادگیری جمعی (Ensemble) می‌باشد. منظور از روش دسته‌بندی یا Bagging میانگین‌گیری از نتیجه پیش‌بینی چندین روش طبقه‌بندی است تا دقت پیش‌بینی‌ها افزایش یابد. در این روش ایستگاه‌ها دسته‌بندی نمی‌شوند بلکه نتایج عددی روش‌های M5P، RT و REPT با هم ترکیب می‌شوند (Breiman, 1996).

معیارهای ارزیابی

برای مقایسه مقادیر بارش روزانه تبریز با مقادیر به‌دست آمده از مدل‌های مورد مطالعه از شاخص‌های ضریب همبستگی (R)، ریشه میانگین مربعات خطا^۲ (RMSE)، ضریب نش ساتنکلیف^۳ (NS)، ضریب ویلموت اصلاح شده^۴ (Dr) و میانگین خطای مطلق^۵ (MAE) استفاده شد. بهترین مدل مربوط به مدلی با RMSE کمتر و R، NS و Dr نزدیک به یک خواهد بود. فرمول‌های آماره‌های فوق به ترتیب در روابط ۱ تا ۵ ارائه گردیده است.

$$R = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \cdot \sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_i - Y_i)^2}{N}} \quad (2)$$

$$NS = 1 - \left(\frac{\sum_{i=1}^N (X_i - Y_i)^2}{\sum_{i=1}^N (X_i - \bar{X})^2} \right) \quad (3)$$

$$D_i = \begin{cases} \frac{C \sum_{i=1}^N |X_i - \bar{X}|}{\sum_{i=1}^N |Y_i - X_i|} - 1 & \text{when } \sum_{i=1}^N |Y_i - X_i| > C \sum_{i=1}^N |X_i - \bar{X}| \\ 1 - \frac{\sum_{i=1}^N |Y_i - X_i|}{C \sum_{i=1}^N |X_i - \bar{X}|} & \text{when } \sum_{i=1}^N |Y_i - X_i| \leq C \sum_{i=1}^N |X_i - \bar{X}| \end{cases} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^N |Y_i - X_i| \quad (5)$$

در روابط فوق، Y_i مقدار بارش محاسبه شده از مدل، X_i بارش مشاهداتی، N تعداد داده‌ها و C ضریبی است که از رابطه ۶ محاسبه می‌شود.

الگوریتم موجود در یادگیری ماشینی بوده که درختان تک مدل با ایده‌های جنگل تصادفی ترکیب شده‌اند. درختان نمونه، درختان تصمیم‌گیری هستند که در آن هر برگ یک مدل خطی دارد که برای زیرفضای محلی توصیف شده توسط این برگ بهینه شده است. جنگل‌های تصادفی نشان داده‌اند که عملکرد درختان تصمیم واحد را به میزان قابل توجهی بهبود می‌بخشد (Kalmegh 2015).

مدل درختی M5P

الگوریتم M5P بازسازی منطقی و توسعه‌یافته از M5 است که توسط ونگ و ویتن در سال ۱۹۹۷ معرفی شده است. مدل درختی M5P قابلیت پیش‌بینی متغیرهای پیوسته عددی از روی صفات عددی را دارد و نتایج پیش‌بینی شده به‌صورت مدل‌های رگرسیونی خطی چندگانه در برگ‌های درخت ظاهر می‌شوند. معیار تقسیم در یک گره بر اساس انتخاب انحراف معیار مقادیر خروجی که به آن گره می‌رسند به‌عنوان معیاری از خطا است. با آزمون هر صفت (پارامتر) در گره کاهش مورد انتظار در خطا محاسبه می‌شود (Wang and Witten, 1997).

کاهش خطای هرس درخت

REPT از منطق درخت رگرسیون استفاده می‌کند و چندین درخت را در تکرارهای مختلف ایجاد می‌کند. پس از آن بهترین درخت را از بین همه درختان تولید شده انتخاب می‌کند و به‌عنوان نماینده در نظر می‌گیرد. کاهش خطا هرس درخت، یادگیری سریع درخت تصمیم‌گیری است و بر اساس کسب اطلاعات یا کاهش واریانس، یک درخت تصمیم می‌سازد. REP فقط یک بار مقادیر صفات عددی را مرتب می‌کند (Kalmegh, 2015).

دسته‌بندی

روش دسته‌بندی که برای اولین بار توسط بریمن در سال ۱۹۹۶ ارائه شد، چندین یادگیر پایه را به صورت موازی به هم متصل می‌کند تا واریانس مجموعه را کاهش دهد. هر یادگیر پایه با استفاده از الگوریتم یادگیری یکسان بر روی یک نسخه بوت استرپ^۱ آموزش می‌بیند، سپس خروجی‌های این یادگیرندگان پایه با رأی اکثریت (برای طبقه‌بندی) یا میانگین‌گیری (برای رگرسیون) برای به‌دست آوردن خروجی نهایی جمع می‌شوند. برای دستیابی به عملکرد بهتر و قوی‌تر، یادگیرندگان پایه در یک گروه باید دقیق و متنوع باشند. در ادبیات علم داده و داده‌کاوی روش دسته‌بندی یا Bagging یکی از

4- Refined Index of Agreement

5- Mean Absolute Error

1- Bootstrap

2- Root Mean Square Error

3- Nash Sutcliffe

داده‌های اصلی بارش تبریز به همراه اجزای تجزیه شده آن در شکل ۳ نشان داده شد.

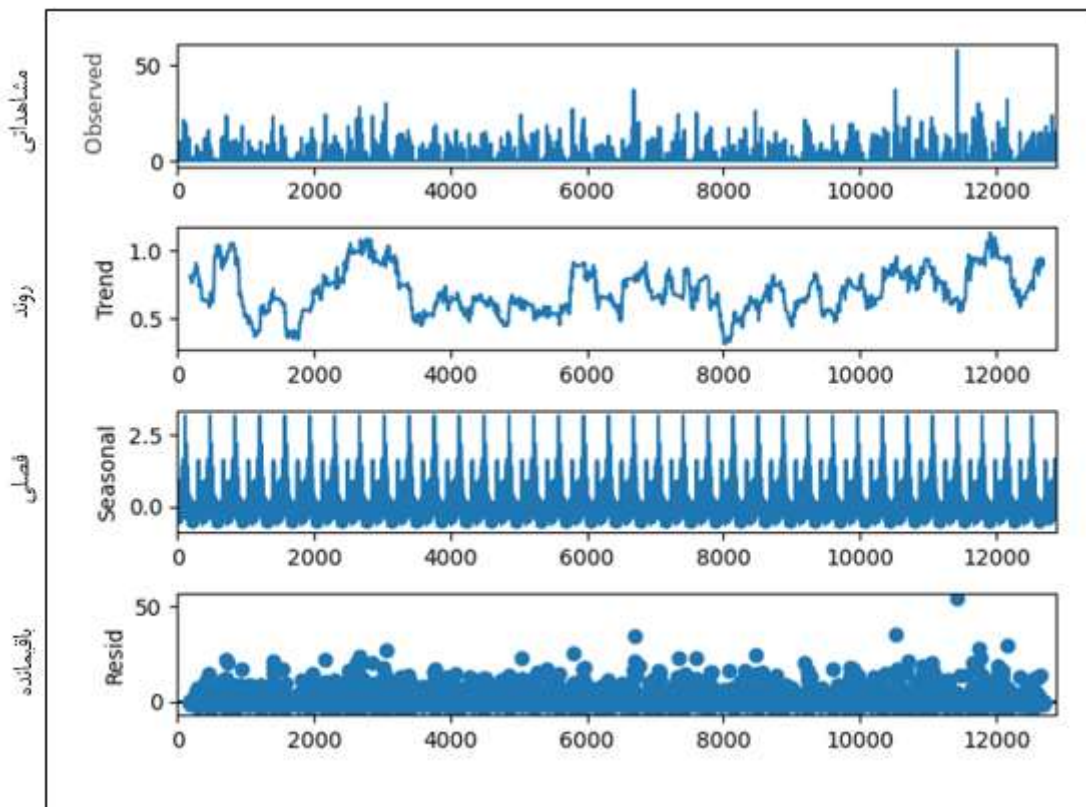
در ابتدا، نتایج به‌دست‌آمده از ۶ سناریوی ورودی مورد استفاده در روش‌های مبتنی بر درخت به همراه روش دسته‌بندی در جدول ۴ نشان ارائه گردید.

با توجه به نتایج جدول ۴ روش‌های RT و REPT وقتی به‌عنوان الگوریتم پایه روش دسته‌بندی قرار گرفتند دقت مدل افزایش و خطای آن کاهش یافت. اما در مورد روش MSP این‌گونه نبود و نتایج در حد کمی ضعیف‌تر شدند. از دیگر نتایجی که از جدول فوق گرفته شد این بود که بارش تبریز تا حد زیادی تحت تأثیر بارش سه‌هنگ است زیرا در اکثر مدل‌ها نتایج سناریو اول که شامل ایستگاه سه‌هنگ است، نتایج قابل قبولی را ارائه کرده است. یکی از دلایل این دقت بالا می‌تواند همبستگی زیاد بارش تبریز با سه‌هنگ باشد. هدف از به کار بردن روش‌های داده‌کاوی، آسان بودن استفاده از آن‌ها و به کار بردن پارامترهای کمتر به‌عنوان ورودی مدل است. برای این‌که از بین تمام روش‌های استفاده شده، سناریو برتر انتخاب شود از نمودار راداری (شکل ۴) استفاده شد.

$$\frac{1}{C} = \frac{MAE}{MAD} = \frac{\frac{1}{N} \sum_{i=1}^N |X_i - Y_i|}{\frac{\sum_{i=1}^N |X_i - \bar{X}|}{N}} \quad (6)$$

نتایج و بحث

در پژوهش حاضر برای برآورد میزان بارش روزانه تبریز در دوره آماری ۳۶ ساله از بارش‌های ایستگاه‌های حوضه دریاچه ارومیه شامل سه‌هنگ، سراب، ارومیه، مراغه و مهاباد در ترکیب‌های ورودی مختلف استفاده شد. مدل‌سازی بارش با استفاده از روش‌های داده‌کاوی مبتنی بر درخت از جمله REPT و RT، MSP انجام شد. سپس تمام روش‌های استفاده شده در این مطالعه به‌عنوان الگوریتم پایه روش دسته‌بندی در نظر گرفته شد تا کارایی این روش سنجیده شود. در جهت بهبود نتایج مدل‌سازی به بررسی تأثیر مؤلفه‌های تجزیه بارش تبریز در ترکیب‌های ورودی مختلف پرداخته شد. نمودار سری زمانی

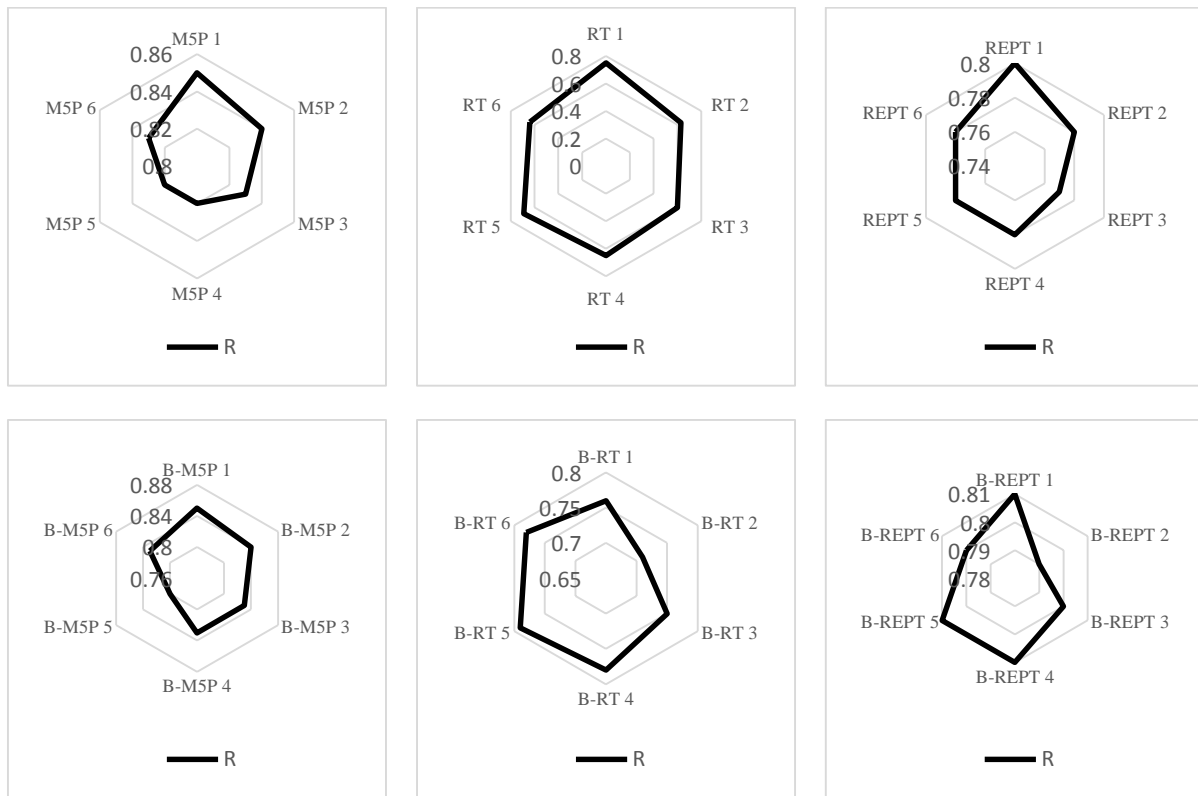


شکل ۳- نمودار سری زمانی بارش روزانه تبریز به همراه مؤلفه‌های تجزیه شده آن
Figure 3- Tabriz daily rainfall time series diagram with its components

جدول ۴- معیارهای ارزیابی روش‌های مبتنی بر درخت و روش دسته‌بندی براساس سناریوهای تعریف شده بدون مؤلفه‌های تجزیه فصلی-روند

Table 4- Evaluation criteria for tree-based and bagging methods based on defined Scenarios without seasonal analysis components - trend

Scenario	روش (Method)														
	M5P					RT					REPT				
	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE
1	0.85	1.54	0.69	0.79	0.58	0.75	1.82	0.56	0.75	0.63	0.80	1.70	0.62	0.78	0.59
2	0.84	1.56	0.68	0.79	0.57	0.63	2.23	0.35	0.69	0.70	0.78	1.75	0.59	0.80	0.56
3	0.83	1.58	0.67	0.80	0.55	0.60	2.34	0.28	0.69	0.70	0.77	1.77	0.59	0.79	0.57
4	0.82	1.59	0.67	0.79	0.57	0.65	2.16	0.39	0.72	0.67	0.78	1.72	0.61	0.80	0.56
5	0.82	1.60	0.66	0.79	0.57	0.69	2.07	0.44	0.71	0.68	0.78	1.72	0.61	0.80	0.56
6	0.83	1.58	0.67	0.79	0.57	0.64	2.20	0.36	0.70	0.68	0.78	1.74	0.60	0.80	0.55
Scenario	B-M5P					B-RT					B-REPT				
	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE
	1	0.85	1.55	0.68	0.78	0.58	0.76	1.81	0.57	0.75	0.63	0.81	1.67	0.63	0.77
2	0.84	1.57	0.68	0.78	0.58	0.71	1.95	0.50	0.74	0.64	0.79	1.72	0.61	0.79	0.57
3	0.83	1.60	0.66	0.78	0.59	0.75	1.82	0.56	0.79	0.58	0.80	1.68	0.63	0.81	0.54
4	0.83	1.58	0.67	0.79	0.57	0.78	1.73	0.61	0.80	0.55	0.81	1.64	0.64	0.82	0.53
5	0.80	1.67	0.63	0.58	0.58	0.79	1.68	0.63	0.81	0.55	0.81	1.65	0.64	0.82	0.53
6	0.83	1.57	0.67	0.79	0.57	0.78	1.72	0.61	0.80	0.56	0.80	1.67	0.63	0.81	0.54



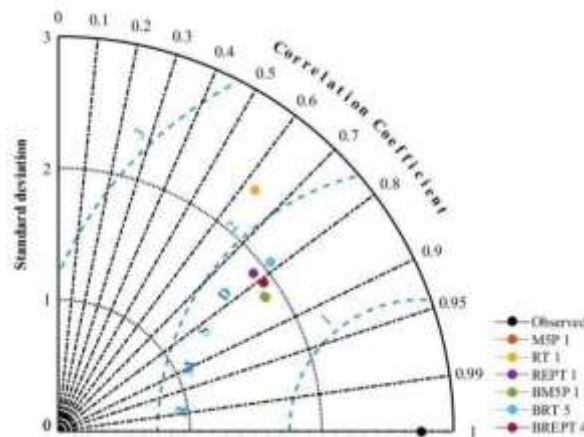
شکل ۴- نمودار راداری ضریب همبستگی برای روش‌های مورد مطالعه
Figure 4- Radar diagram of correlation coefficient for the studied methods

پنجم (سهند، سراب، ارومیه، مراغه و مهاباد) به‌عنوان سناریوهای برتر شناخته شدند که بقیه معیارهای ارزیابی نیز بیانگر برتر بودن این سناریوها بود. در روش B-REPT با توجه به یکسان بودن ضریب همبستگی، معیارهای RMSE، MAE، Dr و NS در نظر گرفته شده و سناریو چهارم (سهند، سراب، ارومیه و مهاباد) به‌عنوان سناریو برتر

نمودار راداری، ۶ ضلعی مربوط به همبستگی بین بارش مشاهداتی و محاسبه شده از مدل‌ها را نشان می‌دهد. نزدیک بودن رأس‌های ۶ ضلعی رسم شده به رأس‌های ۶ ضلعی بیرونی نشان‌دهنده بالا بودن مقدار همبستگی است. بنابراین نتیجه گرفته شد که در روش M5P، RT، REPT و B-M5P سناریو اول (سهند)، در روش B-RT سناریو

انحراف استاندارد با یکدیگر مطابقت دارند ارائه دهد. با توجه به **شکل ۵** بهترین عملکرد، مربوط به سناریو ۱ مدل درخت تصمیم M5P و روش دسته‌بندی با الگوریتم پایه M5P با اختلاف بسیار کم بود.

انتخاب شد. برای داشتن تصور بصری بهتری از نتایج به دست آمده از مدل‌سازی، دیاگرام تیلور برای سناریوهای برتر ترسیم شد (**شکل ۵**). دیاگرام تیلور نموداری دو بعدی است که سه آماره روی آن نشان داده می‌شود. این دیاگرام می‌تواند یک خلاصه آماری مختصر از این که چگونه مدل‌ها از نظر همبستگی، انحراف ریشه میانگین مربع و



شکل ۵- دیاگرام تیلور سناریوهای برتر قبل از تجزیه

Figure 5- Taylor diagram of top scenarios before analysis

جدول ۵- معیارهای ارزیابی روش‌های مبتنی بر درخت و روش دسته‌بندی براساس سناریوهای تعریف شده با در نظر گرفتن مؤلفه‌های تجزیه فصلی-روند

Table 5- Evaluation criteria for tree-based methods and bagging methods based on defined Scenarios considering seasonal analysis components – trend

Scenario	روش (Method)														
	M5P					RT					REPT				
	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE
1	0.97	0.72	0.93	0.98	0.16	0.95	0.89	0.90	0.97	0.22	0.89	1.25	0.79	0.97	
2	0.97	0.72	0.93	0.98	0.16	0.95	0.90	0.89	0.97	0.22	0.94	0.92	0.89	0.98	0.20
3	0.97	0.72	0.93	0.98	0.16	0.95	0.91	0.89	0.97	0.21	0.94	0.91	0.89	0.98	0.20
4	0.97	0.71	0.93	0.98	0.16	0.95	0.90	0.89	0.97	0.22	0.94	0.92	0.89	0.98	0.20
5	0.97	0.71	0.93	0.98	0.16	0.94	0.96	0.88	0.97	0.22	0.94	0.91	0.89	0.98	0.20
6	0.96	0.75	0.93	0.99	0.11	0.95	0.88	0.90	0.98	0.18	0.95	0.89	0.90	0.99	0.15
7	0.96	0.74	0.93	1.00	0.08	0.95	0.86	0.90	0.98	0.16	0.95	0.89	0.90	0.99	0.14
8	0.96	0.80	0.92	0.98	0.17	0.94	0.92	0.89	0.97	0.22	0.94	0.92	0.89	0.97	0.20
9	0.96	0.80	0.92	0.98	0.16	0.95	0.89	0.90	0.97	0.21	0.94	0.92	0.89	0.97	0.20
10	0.96	0.74	0.93	1.00	0.08	0.95	0.85	0.91	0.98	0.16	0.95	0.89	0.90	0.99	0.14
11	0.96	0.74	0.93	1.00	0.08	0.94	0.95	0.88	0.98	0.17	0.95	0.89	0.90	0.99	0.14
12	0.96	0.74	0.93	1.00	0.08	0.95	0.92	0.89	0.98	0.18	0.95	0.89	0.90	0.99	0.14
Scenario	B-M5P					B-RT					B-REPT				
	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE	R	RMSE	NS	Dr	MAE
1	0.98	0.61	0.95	0.98	0.16	0.95	0.88	0.90	0.97	0.20	0.95	0.90	0.89	0.98	0.19
2	0.94	0.93	0.89	0.98	0.17	0.95	0.86	0.90	0.98	0.19	0.95	0.89	0.89	0.98	0.19
3	0.94	0.93	0.89	0.98	0.17	0.95	0.87	0.90	0.98	0.19	0.95	0.90	0.89	0.98	0.19
4	0.95	0.88	0.90	0.98	0.17	0.95	0.86	0.90	0.98	0.18	0.95	0.90	0.89	0.98	0.19
5	0.95	0.88	0.90	0.98	0.17	0.96	0.84	0.91	0.98	0.18	0.95	0.90	0.89	0.98	0.19
6	0.96	0.74	0.93	0.99	0.11	0.96	0.84	0.91	0.99	0.15	0.95	0.88	0.90	0.99	0.14
7	0.96	0.74	0.93	1.00	0.08	0.96	0.86	0.90	0.99	0.15	0.95	0.88	0.90	0.99	0.14
8	0.96	0.82	0.91	0.98	0.17	0.96	0.82	0.91	0.98	0.18	0.95	0.89	0.89	0.98	0.19
9	0.96	0.81	0.91	0.98	0.17	0.95	0.88	0.90	0.98	0.19	0.95	0.89	0.89	0.98	0.19
10	0.96	0.74	0.93	1.00	0.08	0.96	0.84	0.91	0.99	0.13	0.95	0.87	0.90	0.99	0.14
11	0.96	0.74	0.93	1.00	0.08	0.95	0.85	0.90	0.99	0.14	0.95	0.87	0.90	0.99	0.14
12	0.96	0.74	0.93	1.00	0.08	0.95	0.88	0.90	0.98	0.16	0.95	0.88	0.90	0.99	0.14

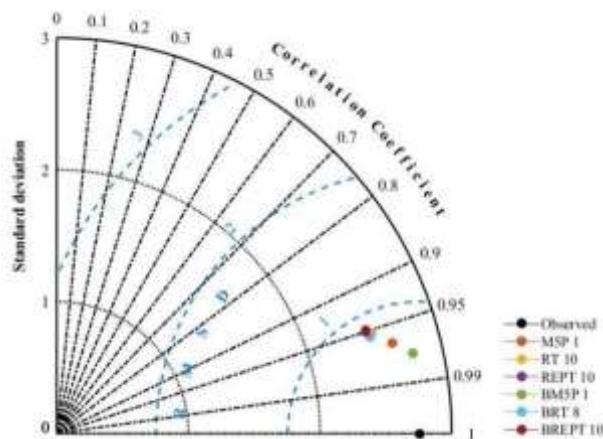
پارامترهای ورودی مد نظر قرار گرفت چرا که کم بودن تعداد پارامتر ورودی باعث ساده و قابل فهم بودن مدل می‌شود (شکل ۶).

با توجه به شکل ۶ در روش‌های M5P و B-M5P سناریو اول (سهند و باقیمانده)، در روش‌های RT، REPT و B-REPT سناریو دهم (باقیمانده، روند، فصلی، سهند و سراب) و در روش B-RT سناریو هشتم (باقیمانده، روند و سهند) به‌عنوان سناریوهای برتر انتخاب شدند. در نتیجه از بین ایستگاه‌های مورد مطالعه ایستگاه سهند به خاطر نزدیکی و همبستگی زیاد و سراب به دلیل همبستگی بیشتر بر روی بارش تبریز تأثیر زیادی می‌گذارند. در حالت کلی روش دسته‌بندی با الگوریتم پایه M5P (B-M5P) در سناریو اول بهترین عملکرد را داشته است. پس می‌توان نتیجه گرفت که هم اضافه کردن مؤلفه‌های تجزیه بارش و هم استفاده از روش دسته‌بندی، نتایج مدل‌سازی با روش‌های داده‌مبنای مبتنی بر درخت را بهبود داده است. در شکل ۷ نمودار سری زمانی بهترین روش (B-M5P1) نشان داده شده است.

زیرا مقدار انحراف استاندارد داده‌های مشاهداتی و انحراف استاندارد داده‌های برآورد شده از مدل، نزدیک به هم بوده و نتایج قابل قبولی را ارائه کرده است. به‌صورت کلی می‌توان نتیجه گرفت که استفاده از روش دسته‌بندی باعث افزایش دقت پیش‌بینی‌ها می‌گردد.

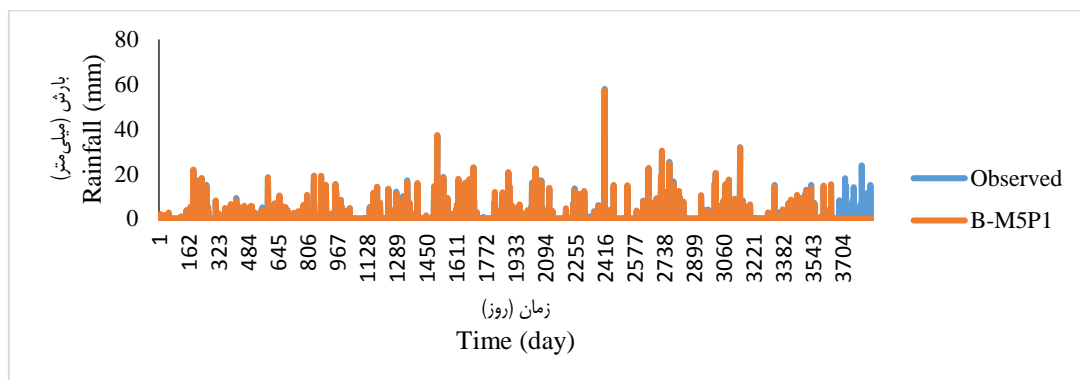
طبق ترکیب‌های ورودی جدول ۳، مدل‌سازی با در نظر گرفتن مؤلفه‌های تجزیه در کنار ایستگاه‌های مورد مطالعه انجام شده و معیارهای ارزیابی در جدول ۵ ارائه شد.

در این مطالعه نتایج مدل‌سازی بدون در نظر گرفتن مؤلفه‌های تجزیه با مدل‌سازی همراه با مؤلفه‌های تجزیه با هم مقایسه شدند. نتایج حاکی از آن بود که اضافه شدن مؤلفه‌های فصلی، روند و باقیمانده به ترکیب‌های ورودی مدل‌سازی باعث بهبود چشمگیر نتایج می‌شود. همچنین استفاده از روش دسته‌بندی نیز در اکثر موارد دقت مدل‌سازی را بیشتر کرد. با در نظر گرفتن تمام معیارهای ارزیابی، سناریوهای برتر مدل‌سازی‌های فوق نیز انتخاب و روی نمودار تیلور نشان داده شد. در صورت مشابه بودن معیارهای ارزیابی، تعداد



شکل ۶- دیاگرام تیلور سناریوهای برتر بعد از تجزیه

Figure 6- Taylor diagram of top scenarios after analysis



شکل ۷- نمودار سری زمانی روش B-M5P1

Figure 7- Time series diagram of B-M5P1 method

روش درخت تصمیم M5P و B-M5P با نتایج نزدیک به هم به‌عنوان دقیق‌ترین مدل معرفی شدند. پس از اضافه کردن مؤلفه‌های روند، فصلی و باقیمانده به ترکیب‌های ورودی، دقت مدل‌ها به‌صورت چشمگیری افزایش یافت و روش B-M5P1 شامل باقیمانده و بارش سه‌سند به‌عنوان سناریو و روش برتر انتخاب شد. در حالت کلی نتیجه گرفته شد که استفاده از روش دسته‌بندی و مؤلفه‌های تجزیه بارش باعث بهبود نتایج مدل‌سازی می‌شود. بارش ایستگاه سه‌سند به‌عنوان ایستگاهی با نزدیک‌ترین فاصله از تبریز و دارای بیشترین همبستگی با بارش تبریز یکی از مؤثرترین پارامترهای ورودی در مدل‌سازی بود. بقیه روش‌ها و سناریوهای استفاده شده نیز دقت قابل قبولی داشتند، لذا در صورت داشتن اطلاعات بارش سایر ایستگاه‌ها می‌توان از آن روش‌ها و ورودی‌ها نیز بهره جست. با توجه به نتایج مطالعه حاضر پیشنهاد می‌گردد برای برآورد میزان بارش یک ایستگاه از مقادیر بارش نزدیک‌ترین ایستگاه استفاده گردد به دلیل اینکه کم بودن فاصله، در اکثر مواقع باعث مشابهت در شرایط اقلیمی دو ایستگاه گردیده و باعث همبستگی بیشتر بین بارش ایستگاه‌ها می‌شود.

با توجه به شکل ۷ نتیجه گرفته شد که مدل به خوبی توانسته است مقادیر بارش روزانه ایستگاه سینوپتیک تبریز را برآورد کند.

نتیجه گیری

در پژوهش حاضر به برآورد میزان بارش روزانه تبریز در دوره آماری ۳۶ ساله (۱۹۸۶-۲۰۲۱) پرداخته شد. برای مدل‌سازی از مقادیر بارش ایستگاه‌های حوضه دریاچه ارومیه از جمله سه‌سند، سراب، ارومیه، مراغه و مهاباد استفاده شد. روش‌های مورد استفاده در مطالعه حاضر شامل گروه روش‌های داده‌مبنای مبتنی بر درخت M5P، درخت تصادفی، کاهش خطای هرس درخت و روش دسته‌بندی بود. از مؤلفه‌های تجزیه فصلی-روند در کنار ایستگاه‌های مذکور، در ترکیب‌های ورودی مختلف استفاده شد تا تأثیر این مؤلفه‌ها در بهبود نتایج بررسی شود.

قبل از تجزیه مقادیر بارش تبریز و استفاده از این مؤلفه‌ها در مدل‌سازی، در اکثر موارد سناریو اول شامل بارش ایستگاه سه‌سند به‌عنوان سناریو برتر شناسایی شد. از بین روش‌های مورد مطالعه نیز

منابع

- Adnan A., Yolanda A. M., and Natasya F. 2021. A comparison of bagging and boosting on classification data: Case study on rainfall data in Sultan Syarif Kasim II meteorological station in Pekanbaru. *Journal of Physics*. <https://doi.org/10.1088/1742-6596/2049/1/012053>.
- Asakereh H., and Akbarzadeh Y. 2017. Simulation of temperature and precipitation changes of Tabriz Synoptic Station using statistical downscaling and Canesm2 climate change model output. *Journal of Geography and Environmental Hazards* 21: 153-174. <https://doi.org/10.22067/GEO.V6I1>. (In Persian with English abstract)
- Barrera A., Oyedele L., Bilal M., Akinosho T., Delgado J., and Akanbi L. 2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications* 7. <https://doi.org/10.1016/j.mlwa.2021.100204>.
- Breiman L. 1996. Bagging predictors. *Machine Learning* 24: 123-140.
- Bushara N., and Abraham A. 2015. Novel Ensemble Method for Long Term Rainfall Prediction. *International Journal of Computer Information Systems and Industrial Management Applications* 7: 116-130.
- Cabezuelo A. 2022. Prediction of Rainfall in Australia Using Machine Learning. *Information* 13(163). <https://doi.org/10.3390/info13040163>.
- Choubin B., Zehtabian Gh., Azareh A., Rafiei-Sardooi E., Sajedi-Hosseini F., and Kisi O. 2018. Precipitation forecasting using classification and regression trees(CART) model: a comparative study of different approaches. *Environmental Earth Sciences*. <https://doi.org/10.1007/s12665-018-7498-z>.
- Dastourani M. T., Habibipour A., Ekhtesasi M. R., Talebi A., and Mahjoobi J. 2013. Evaluation of the Decision Tree Model in Precipitation Prediction (Case study: Yazd Synoptic Station). *Iran-Water Resources Research* 8(3): 14-27. (In Persian with English abstract)
- Endalie D., Hailea G., and Taye W. 2022. Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia. *Water Supply* 3(22). <https://doi.org/10.2166/ws.2021.391>.
- Kalmegh S. 2015. Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News. *International Journal of Innovative Science, Engineering & Technology* 2: 438-446.
- Kira K., and Rendell L.A. 1992. The feature selection problem: Traditional methods and a new algorithm. *AAAI-92 Proceedings of the tenth national conference on Artificial intelligence* 129-134.
- Liyew M., and Melese H. 2021. Machine learning techniques to predict daily rainfall amount. *Journal of Big Data* 8(153). <https://doi.org/10.1186/s40537-021-00545-4>.
- Mishra N., Soni H., Sharma S., and Upadhyay A. 2017. A Comprehensive survey of data mining techniques on time series data for rainfall prediction. *Journal of ICT Research and Applications* 11(2): 168-184. <https://doi.org/10.5614/itbj.ict.res.appl.2017.11.2.4>.

14. Nagahamulla H., Ratnayake U., and Ratnaweera A. 2014. Selecting most suitable members for neural network ensemble rainfall forecasting model. *Recent Advances on Soft Computing and Data Mining* 591-601. https://doi.org/10.1007/978-3-319-07692-8_56.
15. Omidvar K., and Azhdarpoor M. 2013. Comparison of artificial neural network and HEC-HMS model in assessment- runoff in Herat Azam catchment river. *Geographical Research Quarterly* 4: 139-159. (In Persian)
16. Omidvar K., Shafie Sh., Taghizade Z., and Alipoor M. 2014. Evaluating the efficiency of the decision tree model in predicting rainfall in Kermanshah synoptic station. *Journal of Applied Research in Geographical Sciences* 14(34): 89-110. (In Persian)
17. Sattari M.T., and Nahrein F. 2014. Monthly rainfall prediction using Artificial Neural Networks and M5 model tree (Case study: Station s of Ahar and Jolfa). *Journal of Irrigation and Water Engineering* 4(14): 83-98. (In Persian with English abstract)
18. Sattari M. T., Falsafian K., Irvem A., Shahab S., and Qasem S. 2020. Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall. *Engineering Applications of Computational Fluid Mechanics* 14(1): 1078-1094. <https://doi.org/10.1080/19942060.2020.1803971>.
19. Tahroudi M., Ahmadi F., and Khalili K. 2017. Evaluation the Trend and Trend Chang Point of Urmia Lake Basin Precipitation. *Journal of Water and Soil* 31: 644-659. <https://doi.org/10.22067/JSW.V31I2.55338>. (In Persian with English abstract)
20. Wang Y., and Witten I.H. 1997. "Inducing model trees for continuous classes", in *Proceedings of the Ninth European Conference on Machine Learning*. Prague, Czech Republic: Springer 128-137.
21. Yobero C. 2018. Determining Creditworthiness for Loan Applications Using C5.0 Decision Trees. RPub by RStudio.
22. Yu N., and Haskins T. 2021. Bagging Machine Learning Algorithms: A Generic Computing Framework Based on Machine-Learning Methods for Regional Rainfall Forecasting in Upstate New York. *Informatics* 8 (47). <https://doi.org/10.3390/informatics8030047>.
23. Zhou Z.H. 2012. *Ensemble Methods: Foundations and Algorithms* (New York (US): Chapman & Hall/CRC Press).