

بکارگیری الگوریتم‌های یادگیری ماشین در پیش‌بینی بارش روزانه با استفاده از داده‌های مشاهداتی سطح زمین و جو بالا، مطالعه موردی: شهر مشهد

امیرحسین بابائیان^{*}، مهدی رستم زاده^۱، مصطفی فاضلی^۱

۱- کارشناسی ارشد، گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه فردوسی مشهد، خراسان رضوی.

چکیده

در این پژوهش، عملکرد مدل‌های یادگیری ماشین در پیش‌بینی بارش یک روزه ایستگاه مشهد با استفاده از داده‌های روزانه دوره ۲۰۰۰ تا ۲۰۲۳ مورد بررسی قرار گرفت. برای این منظور، سه گروه (سناریو) از داده‌ها در نظر گرفته شد: (۱) داده‌های سطح زمین، (۲) داده‌های جو بالا و (۳) ترکیب هر دو مجموعه داده. پیش از مدل‌سازی، متغیرهای دارای هم‌خطی بالا با استفاده از شاخص VIF شناسایی و حذف شدند تا از استقلال آماری ویژگی‌های ورودی اطمینان حاصل شود. در سناریوی ترکیبی، پس از ادغام داده‌ها، از روش انتخاب ویژگی حذف شناور ($SFFS$) برای گزینش شش شاخص مؤثر استفاده گردید که شامل چهار متغیر جو بالا (Spe_Hum500 ، Rel_Hum300 ، $v700$ و 7500) و دو متغیر سطح زمین (nm و $Umax$) بودند. نتایج ارزیابی مدل‌ها نشان داد که در تمامی سناریوهای مطالعه‌شده، الگوریتم $CatBoost$ به عنوان بهترین مدل شناسایی شد. در سناریوی اول (داده‌های سطح زمین)، $CatBoost$ با MAE برابر 0.841 ، $RMSE$ معادل 3.309 ، R^2 برابر 0.171 و $Adjusted R^2$ برابر 0.168 ، عملکردی برتر از سایر الگوریتم‌ها داشت. در سناریوی دوم (داده‌های جو بالا)، $CatBoost$ دوباره پیش‌تاز ظاهر شد و با MAE برابر 0.772 ، $RMSE$ معادل 2.294 ، R^2 برابر 0.182 و $Adjusted R^2$ برابر 0.175 ، کارایی بالاتری نسبت به رقبا از خود نشان داد. در نهایت، در سناریوی سوم (ترکیب داده‌های سطح زمین و جو بالا)، $CatBoost$ با بهره‌گیری از شش ویژگی منتخب، بهترین نتایج را کسب کرد: $R^2 = 0.190$ ، $Adjusted R^2$ برابر 0.188 و $RMSE = 2.283$. یافته‌ها نشان می‌دهد که تلفیق داده‌های سطح زمین و جو بالا، همراه با انتخاب بهینه ویژگی‌ها، نه تنها دقت پیش‌بینی را افزایش می‌دهد، بلکه تعادل مناسبی میان پیچیدگی مدل و توان پیش‌بینی را ایجاد می‌کند. مدل $CatBoost$ در این سناریو توانسته است بارش‌های خفیف و متوسط را با خطای کم پیش‌بینی کند و روند کلی نوسانات بارش را به خوبی دنبال نماید.

کلید واژه‌ها: پیش‌بینی روزانه بارش، الگوریتم یادگیری، داده‌های سطح زمین، داده‌های جو بالا، مشهد.

مقدمه

پیش‌بینی بارش یکی از جنبه‌های حیاتی پیش‌بینی وضع هوا است که نقش کلیدی در کشاورزی هوشمند، مدیریت منابع آب و هشدار زودهنگام مخاطرات جوی ایفا می‌کند. هواشناسان در حوزه‌های گوناگون با بهره‌گیری از ترکیبی از علوم پایه و فناوری‌های نوین، و با اتکا به پارامترهایی نظیر رطوبت، سرعت باد و دما، اقدام به پیش‌بینی میزان بارش می‌نمایند (Iliger & Pinto, 2022).

استفاده از تکنیک‌های یادگیری ماشین به استخراج اطلاعات ارزشمند از داده‌های هواشناسی کمک می‌کند و موجب درک عمیق‌تری از پویایی‌های اقلیمی می‌گردد؛ که تصمیم‌گیری آگاهانه در حوزه‌های مختلف مرتبط با خدمات هواشناسی را ممکن می‌سازد. از این رو، پژوهشگران متعددی به بررسی اهمیت یادگیری ماشین، به‌ویژه در زمینه داده‌های هواشناسی، در راستای شناسایی و استخراج الگوهای پنهان و دانش فرآیندهای جوی پرداخته‌اند (Gaikwad & Nikam, 2013). آن‌ها چالش‌های مرتبط با پیش‌بینی دقیق بارش را بررسی کرده و راه‌حلی با استفاده از الگوریتم‌های یادگیری ماشین پیشنهاد داده‌اند.

(Hasan et al., 2016) یک مدل قوی پیش‌بینی بارش با استفاده از رگرسیون بردار پشتیبان (SVR) همراه با توابع کرنل مختلف و تکنیک پیش‌پردازش داده معرفی کردند. سه مدل متمایز برای پیش‌بینی بارش انباشته، شدت بارش روزانه و میانگین بارش پیشنهاد شد که هر یک شامل زیرمدل‌هایی برای پیش‌بینی‌های یک‌روزه، هفت‌روزه و ده‌روزه بودند. نتایج حاصل نشان داد که رویکرد مبتنی بر SVR به‌ویژه در پیش‌بینی‌های هفت‌روزه، دقت بالاتری نسبت به روش‌های مرسوم دارد. همچنین، (C & Sanadi, 2021) عملکرد رگرسیون لجستیک و SVR را در پیش‌بینی بارش در منطقه چیتاگونگ هند، طی سال‌های ۲۰۱۲ تا ۲۰۱۷، مورد ارزیابی قرار دادند. یافته‌های آنان نشان داد که الگوریتم رگرسیون لجستیک با دقت ۹۶.۰۶ درصد عملکرد بهتری نسبت به SVR ارائه می‌دهد.

برخی پژوهش‌ها به شناسایی متغیرها/ویژگی‌های جوی و پیش‌بینی آن‌ها با استفاده از تکنیک‌های یادگیری ماشین پرداخته‌اند. به‌عنوان نمونه، (Liyew & Melese, 2021) بر پیش‌بینی بارش روزانه در منطقه شمال‌غربی اتیوپی تمرکز داشتند؛ ناحیه‌ای که بارش‌ها تأثیر بسزایی بر جریان رود نیل

دارند. در این پژوهش، عملکرد سه الگوریتم یادگیری ماشین شامل رگرسیون خطی چندمتغیره (MLR)، جنگل تصادفی (RF) و تقویت گرادیان شدید (XGBoost) مقایسه شد. نتایج نشان داد که الگوریتم XGBoost با داشتن مقادیر کمتر خطای مطلق میانگین (MAE) و ریشه میانگین مربعات خطا (RMSE)، عملکرد دقیق‌تری نسبت به MLR و RF در پیش‌بینی شدت بارش روزانه ارائه می‌دهد. برای بررسی توسعه روش‌های پیش‌بینی هواشناسی، (T. et al., 2020) بر مدل‌های پیش‌بینی عددی وضع هوا و اهمیت پیشرفت‌های فناوری در سیستم‌های محاسباتی و اندازه‌گیری تمرکز کردند. آزمایش‌ها شامل مقایسه عملکرد جنگل تصادفی، XGBoost و شبکه‌های حافظه بلندمدت کوتاه‌مدت (LSTM) در پیش‌بینی طبقه بارش بر اساس داده‌های رادار بود. نتایج نشان داد که شبکه‌های عصبی، به‌ویژه LSTM، عملکرد بهتری نسبت به الگوریتم‌های جنگل تصادفی و XGBoost داشته است. به‌طور مشابه، (Endalie et al., 2022) با استفاده از شبکه‌های LSTM، مدلی مبتنی بر یادگیری عمیق برای پیش‌بینی بارش روزانه در منطقه جیما در کشور اتیوپی توسعه دادند. داده‌های مورد استفاده در این مطالعه شامل اطلاعات هواشناسی طی سال‌های ۱۹۸۵ تا ۲۰۱۷ بود. عملکرد مدل LSTM با چند روش یادگیری ماشین شامل K-نزدیک‌ترین همسایه (KNN)، درخت تصمیم (DT)، پرسپترون چندلایه (MLP) و ماشین بردار پشتیبان (SVM) مقایسه شد. نتایج حاصل نشان داد که مدل LSTM با دستیابی به RMSE برابر با ۰.۰۱، MAE معادل ۰.۴۷۸۶ و ضریب تعیین (R^2) برابر با ۰.۹۹۷۲، دقت بالاتری نسبت به سایر مدل‌ها ارائه کرده است. در مطالعه‌ای دیگر، (van der Merwe et al., 2023) به بررسی تأثیر دمای سطح دریا (SST) و فاصله از ساحل شرقی در استان مپومالانگا واقع در آفریقای جنوبی بر بارش منطقه‌ای پرداختند. آن‌ها با بهره‌گیری از مدل‌های خطی و جنگل تصادفی، به پیش‌بینی فصلی دمای حداقل و بارش با استفاده از متغیرهای جغرافیایی و اقلیمی پرداختند. نتایج نشان داد که عوامل جغرافیایی نظیر عرض و طول جغرافیایی، ارتفاع از سطح دریا و فاصله از دریا تأثیر قابل‌توجهی بر الگوهای بارش دارند.

برای مقابله با چالش‌های کمبود آب و کاهش تأثیر آن بر تولید برق آبی، (Dash et al., 2018) به پیش‌بینی بارش در کرالا، ایالتی از هند که به پیش‌بینی دقیق بارش برای اقتصاد و معیشت

et al., 2022) پنج مدل شامل LSTM، LSTM پشته‌ای (Bidirectional-) LSTM، دوسویه LSTM، XGBoost و یک مدل ترکیبی را برای پیش‌بینی بارش ساعتی در پنج شهر انگلستان، با استفاده از داده‌های اقلیمی سال‌های ۲۰۰۰ تا ۲۰۲۰، مورد مقایسه قرار دادند. ارزیابی‌ها بر اساس معیارهای Loss، RMSE و MAE انجام شد. نتایج نشان داد که مدل‌های Bidirectional-LSTM و Stacked-LSTM با دو لایه پنهان عملکرد بهتری نسبت به سایر مدل‌ها داشته و در پیش‌بینی بارش ساعتی موفق‌تر عمل کرده‌اند.

در مطالعه‌ای جامع، (Li et al., 2024) به تحلیل مقایسه‌ای عملکرد مدل‌های شبکه عصبی بازگشتی (RNN)، واحدهای بازگشتی کنترل شده (GRU) و LSTM در پیش‌بینی بارش در سه شهر اندونزی پرداختند. این تحقیق بر پایه ۱۲۸۵۲ داده هواشناسی در بازه زمانی ۱۷۹۷ تا ۲۰۱۴ انجام شد. ارزیابی بر اساس معیارهای RMSE، MAE و R^2 صورت گرفت و نتایج نشان داد که مدل LSTM با مقادیر RMSE برابر با ۰.۱۲۸۹ و R^2 معادل ۰.۹۹۹۵ عملکرد برتری نسبت به سایر مدل‌ها داشته و در شناسایی الگوهای بلندمدت جوی و پیش‌بینی دقیق بارش مؤثرتر عمل کرده است.

مطالعه‌ای توسط (Qorban Mahtabi, Farshid Taran, 2018) به پیش‌بینی بارش روزانه در شهر اصفهان پرداخت. آن‌ها با استفاده از روش‌های ANN، KNN، SVC و DT، داده‌های هواشناسی روزهای گذشته (۱ تا ۷ روز قبل) را مورد تحلیل قرار دادند. نتایج نشان داد که شرایط بارش روزانه تا دو روز آینده با دقتی معقول (۷۸٪) قابل پیش‌بینی است. این پژوهش همچنین مشخص کرد که هرچه زمان پیش‌بینی به روز وقوع نزدیک‌تر باشد، استفاده از پارامتر رطوبت نسبی باعث افزایش دقت پیش‌بینی می‌شود. در حالی که عملکرد هر چهار مدل مشابه بود، مدل DT به دلیل سادگی، غیرپارامتریک بودن و ارتباط مستقیم با متغیرهای ورودی، از نظر کاربردی برجسته‌تر ظاهر شد و ابزاری مناسب برای شبیه‌سازی الگوهای بارش محسوب شد. در همین راستا، (Mohammadreza, Fallahi, Hadi Varwani, 2011) از داده‌های روزانه دما، فشار، ساعت آفتابی و رطوبت نسبی برای سه ایستگاه هواشناسی در بندر انزلی، اراک و قم در طول یک سال برای پیش‌بینی بارش استفاده کردند. آن‌ها اثربخشی مدل DT را، به‌ویژه در حوضه‌های کوچک و کوهستانی با زمان تمرکز کم،

خود وابسته است، پرداختند. آن‌ها از مدل‌های هوش مصنوعی، شامل KNN، شبکه عصبی مصنوعی (ANN) و ماشین یادگیری شدید (ELM)، برای پیش‌بینی بارش‌های موسمی تابستانی و پسموسمی استفاده کردند. این الگوریتم‌ها با استفاده از MAE، RMSE و خطای مطلق میانگین مربعات (MASE) ارزیابی شدند و الگوریتم ELM نتایج دقیق‌تری نسبت به دیگران ارائه داد. در همین راستا، (Mahajan & Sharma, 2022) نیز یافته‌هایی ارائه کردند که می‌تواند در برنامه‌ریزی کشاورزی و مدیریت منابع آبی مؤثر باشد. آن‌ها از مدل‌های SVR، MLR و رگرسیون LASSO بهره بردند. نتایج نشان داد که مدل SVR، به‌ویژه در ثبت روابط غیرخطی موجود در داده‌ها، عملکرد بهتری نسبت به MLR دارد. بر اساس معیار MAE، مدل بهینه‌شده SVR بهترین عملکرد پیش‌بینی را نشان داد. در راستای ارتقاء مدیریت منابع آبی با تمرکز ویژه بر پیش‌بینی سیلاب، (Ridwan et al., 2021) به پیش‌بینی بارش در دریاچه کنیور واقع در منطقه هولو ترنگانو، مالزی پرداختند. در این تحقیق، از چهار الگوریتم شامل رگرسیون درخت تصمیم تقویت‌شده (BDTR)، رگرسیون جنگل تصمیم (DFR)، رگرسیون خطی بیزی (BLR) و رگرسیون شبکه عصبی (NNR) استفاده شد. این مطالعه دو رویکرد پیش‌بینی را دنبال کرد: نخست، استفاده از تابع خود همبستگی (ACF) بر پایه داده‌های تاریخی و دوم، استفاده از خطای پیش‌بینی‌شده (PE) مبتنی بر سناریوهای تغییر اقلیم. تحلیل ACF شامل بازه‌های زمانی روزانه، هفتگی، ده‌روزه و ماهانه بود. یافته‌ها نشان داد که مدل BDTR همراه با نرمال‌سازی لگ-نرمال، بهترین نتایج را در بازتولید بارش واقعی و مقادیر PE ارائه کرده و می‌تواند بینش‌های ارزشمندی برای بهبود مدیریت منابع آبی و سازگاری با تغییر اقلیم فراهم کند. در مطالعه‌ای دیگر، (Nguyen et al., 2021) با توسعه یک شبکه عصبی پس‌انتشار، به پیش‌بینی بارش روزانه و ماهانه با استفاده از پارامترهای هواشناسی نظیر دما، نقطه شبنم، رطوبت، فشار، دید افقی و سرعت باد پرداختند. نتایج نشان داد که مدل توسعه‌یافته عملکرد بهتری در پیش‌بینی ماهانه دارد، به‌گونه‌ای که مقادیر ضریب همبستگی، RMSE و MAE برای پیش‌بینی ماهانه به ترتیب ۰.۸۰، ۰.۰۷ و ۰.۰۶ و برای پیش‌بینی روزانه به ترتیب ۰.۸۰، ۰.۲۴ و ۰.۰۹ گزارش شد. این یافته‌ها حاکی از قابلیت اطمینان بالای مدل مذکور برای پیش‌بینی بارش هستند. همچنین (Barrera-Animas

مواد و روش‌ها

داده‌ها

در این پژوهش، به منظور پیش‌بینی بارش روزانه در ایستگاه هواشناسی مشهد، از دو منبع داده‌ای متفاوت شامل مشاهدات سطح زمین، جو بالا و ترکیب آنها استفاده شد. این رویکرد با هدف بهره‌گیری هم‌زمان از شرایط هواشناسی محلی و متغیرهای دینامیکی و ترمودینامیکی بزرگ‌مقیاس جو طراحی شده است. داده‌ها در بازه زمانی سال‌های ۲۰۰۰ تا ۲۰۲۳ گردآوری شده‌اند و جهت ارزیابی عملکرد مدل‌ها در بازه‌های زمانی مجزا، به دو بخش آموزش در بازه زمانی ۲۰۰۰ تا ۲۰۱۷ و آزمون در بازه ۲۰۱۸ تا ۲۰۲۳ تقسیم شده‌اند.

منبع نخست داده‌ها، شامل اطلاعات هواشناسی سطح زمین ثبت‌شده در ایستگاه سینوپتیک مشهد است که توسط سازمان هواشناسی جمهوری اسلامی ایران (IRIMO) گردآوری شده‌اند. متغیرهای موجود در این مجموعه شامل: دمای پیشینه، دمای کمینه، دمای میانگین، رطوبت نسبی پیشینه، رطوبت نسبی کمینه، رطوبت نسبی میانگین، سرعت باد، فشار هوا در سطح دریا و ساعات آفتابی هستند. متغیر هدف در این پژوهش، مقدار بارش روزانه است و تمامی متغیرهای ورودی با یک روز تأخیر زمانی نسبت به روز هدف پیش‌بینی ($lag=1$) در مدل مورد استفاده قرار گرفته‌اند. در جدول ۱، آمار توصیفی متغیرهای سطح زمین شامل میانگین، کمینه، پیشینه، چارک‌های اول تا سوم و انحراف معیار آنها ارائه شده است که به درک بهتر دامنه و توزیع داده‌ها کمک می‌کند.

منبع دوم داده‌ها از پایگاه بازتحلیلی ERA5 وابسته به مرکز پیش‌بینی میان‌مدت هواشناسی اروپا (ECMWF) استخراج شده است. این داده‌ها برای مختصات جغرافیایی ایستگاه مشهد و به صورت روزانه اخذ شدند. متغیرهای استخراج‌شده شامل: ارتفاع ژئوپتانسیل، رطوبت نسبی، رطوبت ویژه، مؤلفه‌های افقی باد (u و v)، چرخندگی و دما در سه تراز فشاری ۷۰۰، ۵۰۰ و ۳۰۰ هکتوپاسکال، به ترتیب به عنوان نماینده تراز زیرین، میانی و فوقانی جو، در نظر گرفته شدند. این سطوح فشاری به گونه‌ای انتخاب شده‌اند که بتوانند لایه‌های کلیدی تروپوسفر را پوشش دهند، چرا که این لایه‌ها تأثیر بسزایی در فرآیندهای جوی منجر به بارش دارند. مشابه داده‌های سطح زمین، در این مجموعه نیز از تمامی متغیرها با یک روز تأخیر زمانی برای پیش‌بینی بارش یک روزه (فردا) استفاده شده است. در جدول ۲، آمار توصیفی متغیرهای جو بالا شامل میانگین،

ارزیابی کردند. نتایج دقت بالای مدل DT را در پیش‌بینی بارش نشان داد.

(Fatemeh Mikaili, 2021) پیش‌بینی بارش روزانه در اردیبهل را با بهره‌گیری از الگوریتم‌های RF و SVR بررسی کرد. داده‌های مورد استفاده شامل بارش، میانگین، حداکثر و حداقل رطوبت نسبی با تأخیر ۱ تا ۳ روز بود که در ترکیب‌های مختلف تحلیل شدند. چهار سناریو بر اساس توابع همبستگی برای کالیبراسیون و اعتبارسنجی مدل‌ها ارزیابی گردید. نتایج نشان داد مدل RF در برآورد دقیق مقادیر پایین و بالا نسبت به SVR عملکرد بهتری دارد. (Rasol Imani, Reza Ghazavi, 2021) نیز روند بارش ماهانه در استان اردبیل را با استفاده از مدل‌های سری زمانی تصادفی خانواده ARIMA پیش‌بینی کردند. بهترین مدل منتخب، کاهش ۳ تا ۱۷ درصدی در بارش سالانه طی چهار سال از پنج سال آینده را نسبت به میانگین ۲۰ سال گذشته پیش‌بینی نمود، به طوری که بیشترین کاهش برای سال ۲۰۲۲ برآورد شد.

همچنین، (Khatere Asghari Tahergorabi, Amir Rajabi Behjat, 2022) دقت مدل‌های پیش‌بینی بارش روزانه را با استفاده از شبکه عصبی Deep ESN و MLR برای ایستگاه‌های هواشناسی هرمزگان مقایسه کردند. نتایج نشان داد که مدل Deep ESN در پیش‌بینی بارش روزانه، بر اساس توابع ارزیابی، از MLR عملکرد بهتری دارد.

در مناطقی که به شدت به کشاورزی وابسته‌اند، مانند ایران، پیش‌بینی بارش نقشی تعیین‌کننده در مدیریت منابع آب، پایداری کشاورزی و کاهش مخاطرات زیست‌محیطی دارد. از آنجا که اقلیم‌های نیمه‌خشک به ویژه نسبت به نوسان‌های بارش حساس هستند، دقت در پیش‌بینی کوتاه‌مدت بارش در این پهنه‌ها اهمیتی دوچندان می‌یابد. با توجه به تمرکز مطالعه‌ی حاضر بر ایران، در این پژوهش با بهره‌گیری از داده‌های هواشناسی روزانه، مدل‌های یادگیری ماشین گروهی برای پیش‌بینی بارش روز بعد (یک روزه) در ایستگاه مشهد توسعه و ارزیابی شده‌اند. هدف اصلی، شناسایی کارآمدترین الگوریتم برای پیش‌بینی کوتاه‌مدت بارش است؛ یافته‌ها نیز می‌توانند بهبود تصمیم‌گیری در مدیریت منابع طبیعی، اجرای راهبردهای کشاورزی هوشمند و طراحی سامانه‌های هشدار زود هنگام بارش‌های سنگین را پشتیبانی کنند.

کمینه، بیشینه، چارک‌های اول تا سوم و انحراف معیار ارائه گردیده است.

جدول ۱- فهرست متغیرهای سطح زمین و توصیف کلی آن‌ها

Variables		Average	Minimum	Maximum	Q1	Q2	Q3	Deviation
Name (unit)	Abbrev.							
Maximum temperature(°C)	Tmax	22.67	-10.10	43.40	14.00	23.70	32.30	10.69
Minimum temperature(°C)	Tmin	9.55	-21.00	29.30	2.35	9.70	17.20	8.60
Average temperature(°C)	Tave	15.86	-14.60	34.80	7.50	16.30	25.00	9.85
Maximum humidity(%)	Umax	66.35	15.00	100.00	43.00	68.00	91.00	24.90
Minimum humidity(%)	Umin	30.45	1.00	100.00	15.00	25.00	41.00	20.59
Average humidity(%)	Uave	47.49	9.88	100.00	27.50	44.12	65.50	22.94
Wind speed(m/s)	WS	2.72	0.00	10.00	1.88	2.62	3.50	1.20
Mean sea level pressure (hpa)	MSLP	1014.66	992.26	1043.72	1007.89	1014.00	1020.90	8.51
Cloudiness	nm	4.38	0.00	8.00	3.15	4.8	6.14	2.10
Rainfall(mm)	R24	0.62	0.00	47.20	0.00	0.00	0.00	2.62

جدول ۲- فهرست متغیرهای جو بالا و توصیف کلی آن‌ها

Variables-level (unit)	Average	Minimum	Maximum	Q1	Q2	Q3	Deviation
GH300(m)	92,385.5	86,826.2	96,350.7	90,693.0	92,477.6	94,172.2	2,062.3
GH500(m)	56,342.3	53,168.8	58,203.1	55,520.7	56,540.1	57,226.5	1,038.9
GH700(m)	30,459.3	28,429.7	31,593.2	30,159.6	30,561.5	30,808.0	461.5
Rel_Hum300(%)	37.6	0.8	107.0	10.9	32.5	60.4	28.8
Rel_Hum500(%)	36.5	0.9	102.8	16.4	33.7	54.0	23.1
Rel_Hum700(%)	44.1	1.6	97.8	28.6	42.0	58.7	20.8
Spe_Hum300(g/kg)	0	0	0.001	0	0	0	0
Spe_Hum500(g/kg)	0.001	0	0.004	0	0.001	0.001	0.001
Spe_Hum700(g/kg)	0.003	0	0.009	0.002	0.003	0.004	0.002
u300(m/s)	21.6	-21.6	78.2	13.5	20.8	28.6	12.1
u500(m/s)	10.5	-13.5	39.7	5.3	10.6	15.5	7.3
u700(m/s)	3.2	-15.8	24.2	-1.2	3.1	7.2	6.1
v300(m/s)	-1.2	-40.7	47.7	-7.5	-1.6	4.8	9.9
v500(m/s)	-1.1	-28.9	27.6	-5.6	-1.7	3.0	6.6
v700(m/s)	-0.6	-15.5	16.6	-3.6	-0.5	2.3	4.3
Vort300(s-1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Vort500(s-1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Vort700(s-1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Temp300(oC)	233.1	216.5	251.7	226.8	231.1	239.9	8.0
Temp500(oC)	258.9	237.9	275.0	253.7	259.0	264.5	6.6
Temp700(oC)	276.4	253.4	292.2	270.9	277.1	282.4	7.1

در نظر گرفته می‌شود. در نهایت، Temp700 دمای هوا در تراز ۷۰۰ هکتوپاسکال را بر حسب کلون بیان می‌کند. این نام‌گذاری در تمامی متغیرها از همین الگو پیروی می‌کند و عدد انتهای هر نام، بیانگر تراز فشاری آن متغیر است. داده‌ها در سه سناریوی مختلف به عنوان ورودی مدل‌ها استفاده شدند؛ در سناریو/گام اول تنها از داده‌های سطح زمین استفاده شده، در گام دوم از داده‌های جو بالا و در گام نهایی، داده‌های سطح زمین و جو بالا با یکدیگر ادغام شده و یک ماتریس ویژگی جامع برای مدل‌سازی تشکیل شده است. در این مرحله، با در نظر گرفتن تأخیر یک‌روزه، اطمینان حاصل

در جدول ۲، متغیر GH500 نشان‌دهنده ارتفاع ژئوپتانسیل در تراز فشاری ۵۰۰ هکتوپاسکال است. متغیرهای Rel_Hum700 و Spe_Hum700 به ترتیب بیانگر رطوبت نسبی (بر حسب درصد) و رطوبت ویژه (بر حسب گرم بر کیلوگرم) در تراز ۷۰۰ هکتوپاسکال هستند. مؤلفه‌های u300 و v300 نیز به ترتیب مؤلفه‌های افقی باد در راستای مداری (غرب به شرق) و نصف‌النهاری (جنوب به شمال) در تراز ۳۰۰ هکتوپاسکال هستند. متغیر Vort500 میزان چرخندگی (بر حسب بر ثانیه) افقی هوا در سطح ۵۰۰ هکتوپاسکال را نشان می‌دهد و به عنوان شاخصی برای بررسی ناپایداری جو

و نقاط ضعف قبلی را جبران می‌کند. این الگوریتم در مواجهه با داده‌های دارای نویز یا الگوهای پیچیده، عملکرد خوبی دارد (Freund & Schapire, 1999).

XGBoost

الگوریتم XGBoost نسخه بهینه شده‌ای از روش گرادینان بوسستینگ است که بسیار سریع و دقیق عمل می‌کند. این الگوریتم در هر مرحله یک درخت جدید می‌سازد که تلاش می‌کند خطاهای قبلی را کاهش دهد. استفاده از منظم‌سازی (Regularization) و توابع زیان جدید باعث می‌شود XGBoost توانایی بالایی در تعمیم‌دهی به داده‌های جدید حتی در شرایط پیچیده داشته باشد (Chen & Guestrin, 2016).

CatBoost

الگوریتم CatBoost نیز نوعی روش گرادینان بوسستینگ است که به ویژه برای داده‌های طبقه‌ای یا متنی بسیار خوب عمل می‌کند. این الگوریتم نیاز زیادی به پیش‌پردازش ندارد و با استفاده از تکنیک‌های خاصی مثل یادگیری ترتیبی، از بیش‌برازش جلوگیری می‌کند. در مسائل پیش‌بینی مانند بارش که داده‌ها می‌توانند پیچیده باشند، CatBoost می‌تواند بدون نیاز به تنظیمات زیاد، عملکرد خوبی داشته باشد (Prokhorenkova et al., 2018).

LightGBM

LightGBM یکی از الگوریتم‌های قدرتمند یادگیری گروهی است که در سال ۲۰۱۷ توسط شرکت مایکروسافت معرفی شده است. این الگوریتم بر پایه درخت‌های تصمیم متوالی ساخته می‌شود؛ یعنی درخت‌های جدید به تدریج خطاهای درخت‌های قبلی را اصلاح می‌کنند. LightGBM با بهره‌گیری از روش رشد مبتنی بر برگ (leaf-wise growth) و استفاده از ساختارهای هیستوگرام محور، سرعت آموزش را افزایش داده و مصرف حافظه را کاهش می‌دهد. به دلیل همین ویژگی‌ها، این الگوریتم برای تحلیل داده‌های حجیم و مسائل بزرگ مقیاس بسیار مناسب است (Bian et al., 2023).

شده که تمامی اطلاعات ورودی مربوط به روز قبل از وقوع بارش هستند. این ترکیب داده‌ای امکان در نظر گرفتن هم‌زمان شرایط سطحی و جو بالا را فراهم کرده و به ارتقای عملکرد مدل‌های پیش‌بینی بارش کمک می‌کند.

الگوریتم‌های یادگیری ماشین

در این پژوهش، برای پیش‌بینی بارش روزانه از مجموعه‌ای متنوع از الگوریتم‌های رگرسیون مبتنی بر یادگیری گروهی استفاده شده است. این الگوریتم‌ها شامل جنگل تصادفی، LightGBM، XGBoost، CatBoost و AdaBoost هستند. الگوریتم‌های یادگیری گروهی با ترکیب چندین درخت تصمیم، توانایی بالایی در شناسایی الگوهای پیچیده، کاهش خطا و جلوگیری از بیش‌برازش دارند و برای مدل‌سازی پدیده‌های غیرخطی مانند بارش گزینه‌های مناسبی محسوب می‌شوند.

لازم به ذکر است که در مراحل اولیه پژوهش، چند مدل پایه دیگر نظیر رگرسیون چندجمله‌ای، ElasticNet، درخت تصمیم (DT)، اسپلاین و MLP نیز پیاده‌سازی و ارزیابی شدند؛ اما به دلیل عملکرد ضعیف‌تر در برآورد بارش نسبت به الگوریتم‌های گروهی، از ارائه جزئیات نتایج آن‌ها صرف‌نظر شده است.

جنگل تصادفی

جنگل تصادفی یکی از الگوریتم‌های قدرتمند در یادگیری گروهی است که از ترکیب چندین درخت تصمیم بهره می‌برد. در این روش، هر درخت روی زیرمجموعه‌ای تصادفی از داده‌ها آموزش می‌بیند و در هر تقسیم‌بندی، تنها بخشی از ویژگی‌ها بررسی می‌شود. این ساختار باعث افزایش دقت، کاهش واریانس و مقاومت در برابر بیش‌برازش می‌شود. جنگل تصادفی در شناسایی روابط پیچیده و الگوهای غیرخطی عملکرد بسیار مناسبی دارد (BREIMAN, 2001).

AdaBoost

الگوریتم AdaBoost یک روش مرحله‌ای است که در آن چند مدل ساده به صورت پشت سر هم آموزش داده می‌شوند. در هر مرحله، مدل جدید تلاش می‌کند خطاهای مدل قبلی را اصلاح کند. به این ترتیب، مدل نهایی به مرور بهتر می‌شود

معیارهای ارزیابی عملکرد مدل‌ها

می‌کند که چند درصد از نوسانات داده‌های واقعی توسط مدل قابل پیش‌بینی است. مقدار این شاخص بین صفر تا یک قرار دارد؛ هرچه مقدار آن به ۱ نزدیک‌تر باشد، نشان‌دهنده برازش بهتر مدل با داده‌ها است. فرمول محاسبه ضریب تعیین به صورت زیر تعریف می‌شود (Sattari et al., 2024):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Adjusted R²

ضریب تعیین تعدیل‌شده، نسخه اصلاح‌شده از ضریب تعیین است که تأثیر تعداد متغیرهای ورودی مدل را نیز در نظر می‌گیرد. این شاخص زمانی اهمیت بیشتری پیدا می‌کند که تعداد ویژگی‌های مدل زیاد باشد، زیرا برخلاف R² که با اضافه شدن هر متغیر جدید حتی اگر مفید نباشد افزایش می‌یابد، Adjusted R² فقط زمانی افزایش پیدا می‌کند که متغیر اضافه‌شده واقعا باعث بهبود مدل شود. این ویژگی باعث می‌شود Adjusted R² ابزار قابل‌اعتمادتری برای مقایسه مدل‌هایی با تعداد متفاوتی از متغیرها باشد، زیرا از بیش‌برازش جلوگیری می‌کند و عملکرد مدل را به شکل واقعی‌تری نمایش می‌دهد. فرمول محاسبه آن به صورت زیر تعریف می‌شود (Pande et al., 2024):

$$Adjusted-R^2 = \frac{(1 - R^2)(n - 1)}{(n - p - 1)}$$

استفاده هم‌زمان از R² و Adjusted R² این امکان را فراهم می‌سازد که علاوه بر سنجش میزان قدرت توضیحی مدل، بررسی کنیم آیا افزودن متغیرهای جدید واقعا مفید بوده یا فقط به صورت ظاهری باعث بهبود شاخص شده است. این موضوع در انتخاب مدل بهینه نقش مهمی ایفا می‌کند.

انتخاب ویژگی‌ها

در این پژوهش، به منظور کاهش ابعاد داده و انتخاب مجموعه‌ای بهینه از متغیرهای ورودی که بیشترین تأثیر را بر پیش‌بینی بارش دارند، از روش SFSS استفاده شده است. این روش یک الگوریتم جستجوی ترتیبی تطبیقی است که با افزودن تدریجی ویژگی‌ها به مدل آغاز می‌شود و در هر مرحله، ویژگی‌ای که بیشترین بهبود را در معیار ارزیابی ایجاد می‌کند، به مجموعه انتخاب‌شده اضافه می‌شود. مزیت اصلی

برای ارزیابی عملکردی مدل‌های رگرسیونی در پیش‌بینی بارش روزانه، از چهار معیار استاندارد شامل MAE، RMSE، R² و ضریب تعیین تعدیل‌شده (Adjusted R²) استفاده شده است. این معیارها از زوایای مختلف به بررسی کیفیت مدل می‌پردازند و در کنار هم تصویری جامع از عملکرد پیش‌بینی ارائه می‌دهند. در تمام فرمول‌های این بخش، منظور از n تعداد نمونه‌ها، y_i مقدار واقعی، \hat{y}_i مقدار پیش‌بینی شده، \bar{y} میانگین مقادیر واقعی و p تعداد متغیرهای مستقل مدل است.

RMSE

ریشه میانگین مربعات خطا معیاری برای سنجش دقت پیش‌بینی مدل است که میزان انحراف پیش‌بینی‌ها از مقادیر واقعی را اندازه‌گیری می‌کند. این شاخص، میانگین توان دوم خطاها را محاسبه کرده و سپس از آن جذر می‌گیرد. از آنجا که توان دوم خطاها در محاسبه لحاظ می‌شود، RMSE نسبت به خطاهای بزرگ حساس‌تر است و به خوبی می‌تواند مدل‌هایی را که در پیش‌بینی‌های پرت ضعف دارند شناسایی کند. فرمول محاسبه‌ی آن به صورت زیر است (Chai & Draxler, 2014; Kisi et al., 2025):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

MAE

میانگین قدرمطلق خطا، میانگین اختلاف مطلق بین مقادیر واقعی و پیش‌بینی‌شده را محاسبه می‌کند و نمایی مستقیم از اندازه خطای متوسط مدل ارائه می‌دهد. برخلاف RMSE، این شاخص نسبت به مقادیر پرت حساسیت کمتری دارد، و بنابراین برای داده‌هایی با نویز یا ناهنجاری‌های زیاد می‌تواند ارزیابی دقیق‌تری از عملکرد مدل ارائه دهد و به صورت زیر تعریف می‌شود (Chai & Draxler, 2014; Kisi et al., 2025):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

R²

ضریب تعیین شاخصی است که نشان می‌دهد مدل رگرسیونی تا چه اندازه توانسته تغییرات متغیر وابسته را بر اساس متغیرهای ورودی توضیح دهد. به عبارت ساده‌تر، R² بیان

(روز بعد) در شهر مشهد مورد ارزیابی قرار گرفته است. به‌منظور دستیابی به بهترین عملکرد هر مدل برازش داده شده، تنظیم دقیق پارامترها با استفاده از روش GridSearchCV انجام شد. در این فرآیند، مجموعه‌ای از مقادیر ممکن برای پارامترهای کلیدی هر مدل تعریف و سپس با اجرای اعتبارسنجی متقابل پنج‌تایی، ترکیب بهینه پارامترها بر اساس بیشترین مقدار معیار R^2 انتخاب گردید. این رویکرد موجب شد مدل‌ها به‌صورت عادلانه و بهینه ارزیابی شوند و از بیش‌برازش جلوگیری گردد. در نهایت، برای سنجش دقت مدل‌های آموزش دیده، بر روی داده‌های آموزش و آزمون، از معیارهای آماری $RMSE$ ، MAE ، R^2 و $Adjusted R^2$ استفاده شده است.

مطالعه حاضر در سه مرحله مجزا پیاده‌سازی شده است. در مرحله نخست، تمامی مدل‌ها با استفاده از ویژگی‌های موجود در مجموعه داده سطح زمین آموزش داده شدند. در مرحله دوم، همان الگوریتم‌ها با استفاده از داده‌های جو بالا آموزش دیدند تا تأثیر اطلاعات لایه‌های بالاتر جو بر دقت پیش‌بینی مورد ارزیابی قرار گیرد. در هر دو مرحله، به‌منظور افزایش پایداری مدل و جلوگیری از اثر هم‌خطی میان متغیرها، شاخص VIF محاسبه و ویژگی‌های دارای هم‌خطی بالا حذف شدند. در مرحله سوم، رویکرد ترکیبی با هدف بهره‌گیری هم‌زمان از داده‌های سطح زمین و جو بالا مورد استفاده قرار گرفت. در این مرحله، پس از ادغام دو مجموعه داده و کنترل هم‌خطی با استفاده از VIF، برای کاهش ابعاد و انتخاب مجموعه‌ای بهینه از متغیرهای مؤثر، الگوریتم SFSS به‌کار گرفته شد. انتخاب ویژگی‌ها بر اساس بیشترین مقدار R^2 در اعتبارسنجی متقابل پنج‌تایی انجام گرفت تا ترکیب نهایی ویژگی‌ها با بیشترین دقت و کمترین پیچیدگی تعیین شود.

برای تسهیل مقایسه نتایج، عملکرد پنج مدل یادگیری ماشین در پیش‌بینی بارش روز بعد در جدول ۳ گردآوری شده است. این جدول، نتایج مدل‌ها را در سه سناریوی داده‌ای مختلف ارائه می‌دهد: سناریوی S از داده‌های سطح زمین استفاده می‌کند، سناریوی U مبتنی بر داده‌های جو بالا است، و سناریوی S&U که ترکیبی از هر دو مجموعه داده را در نظر می‌گیرد. در این جدول، هر مدل با دو ردیف متوالی نمایش

SFSS نسبت به روش انتخاب ترتیبی ساده (SFS) در انعطاف‌پذیری آن است؛ به‌طوری که پس از افزودن هر ویژگی جدید، الگوریتم می‌تواند به‌صورت پویا برخی از ویژگی‌های قبلاً انتخاب‌شده را که اثر آن‌ها در ترکیب جدید کاهش یافته است، حذف نماید. این فرآیند "شناور بودن" امکان جلوگیری از گیر افتادن در بهینه‌های محلی و دستیابی به ترکیب ویژگی‌های مؤثرتر را فراهم می‌کند. استفاده از SFSS به‌ویژه در مسائل با داده‌های چندمنبعی - مانند این مطالعه که ترکیبی از داده‌های سطح زمین و جو بالا مورد استفاده قرار گرفته است - اهمیت دارد، زیرا برخی متغیرها ممکن است به‌تنهایی مفید نباشند اما در ترکیب با سایر ویژگی‌ها تأثیر قابل‌توجهی بر عملکرد مدل داشته باشند. در این پژوهش، اجرای SFSS منجر به کاهش تعداد ویژگی‌های ورودی و در نتیجه ساده‌تر شدن مدل، کاهش خطر بیش‌برازش، و بهبود تفسیرپذیری نتایج گردید.

نرمال‌سازی داده‌ها

به‌منظور هم‌مقیاس‌سازی متغیرهای ورودی و جلوگیری از تأثیر نامتوازن ویژگی‌هایی با دامنه مقادیر بزرگ‌تر بر فرآیند آموزش مدل، در این پژوهش از روش نرمال‌سازی Min-Max استفاده شده است. در این روش، مقادیر هر ویژگی به‌گونه‌ای خطی به بازه صفر تا یک تبدیل می‌شوند. این تبدیل باعث می‌شود تمامی متغیرها وزن یکسانی در فرآیند یادگیری داشته باشند و به بهبود همگرایی الگوریتم‌های یادگیری ماشین کمک کند. فرمول این روش به‌صورت زیر می‌باشد:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

که در آن x مقدار اولیه ویژگی، x_{min} و x_{max} به‌ترتیب کمینه و بیشینه مقدار آن ویژگی در مجموعه داده و x' مقدار نرمال‌شده است.

نتایج و بحث

در این بخش، عملکرد پنج مدل یادگیری ماشین - شامل رگرسیون جنگل تصادفی، AdaBoost، XGBoost، CatBoost و LightGBM - برای پیش‌بینی بارش یک روزه

دست آوردند، در حالی که مدل جنگل تصادفی ضعیف‌ترین عملکرد را در این سناریو داشت. اختلاف اندک میان نتایج آموزش و آزمون در تمامی مدل‌ها، نشان‌دهنده‌ی تنظیم مناسب پارامترها و نبود بیش‌برازش قابل توجه است. بر این اساس، مدل CatBoost در سناریوی داده‌های سطح زمین به‌عنوان کارآمدترین روش شناسایی شد؛ به‌گونه‌ای که بر روی داده‌های آزمون مقادیر $R^2 = 0.171$ ، $Adjusted R^2 = 0.168$ ، $RMSE = 2.309$ و $MAE = 0.841$ را ثبت کرده است، در حالی که AdaBoost با $MAE = 0.767$ کمترین خطای مطلق را در میان روش‌ها به خود اختصاص داده است.

داده‌های جو بالا

در سناریو دوم، مدل‌های یادگیری ماشین بر پایه‌ی داده‌های جو بالا برازش داده شدند. مشابه سناریو اول، پیش از اجرای مدل‌ها، به‌منظور جلوگیری از اثر هم‌خطی میان متغیرها، شاخص VIF برای تمامی ویژگی‌ها محاسبه گردید. نتایج اولیه نشان داد که متغیرهای مربوط به ارتفاع ژئوپتانسیل (GH) و دما (Temp) در ترازهای مختلف جوی (۳۰۰، ۵۰۰ و ۷۰۰ هکتوپاسکال) هم‌خطی بالایی با یکدیگر دارند؛ پدیده‌ای که ناشی از رفتار فیزیکی مشابه این پارامترها در سطوح مختلف فشار است. در نتیجه، برای کاهش افزونگی اطلاعات، از هر گروه تنها یک متغیر به‌عنوان نماینده انتخاب گردید؛ به‌طوری‌که از میان ترازهای مختلف ارتفاع ژئوپتانسیل و دما، به‌ترتیب متغیرهای GH700 و Temp300 در مجموعه‌ی نهایی باقی ماندند. پس از این غربال‌گری، مقادیر VIF سایر ویژگی‌ها در محدوده‌ی قابل قبول قرار گرفت و هم‌خطی میان متغیرها به میزان چشمگیری کاهش یافت. بر اساس این مجموعه‌ی پالایش‌شده، آموزش و ارزیابی مدل‌های جو بالا انجام شد و نتایج عملکرد در جدول ۳ گزارش شده است.

نتایج سناریوی دوم در جدول ۳ نشان می‌دهند که در مدل‌های مبتنی بر داده‌های جو بالا، عملکرد مدل‌ها نسبت به سناریوی سطح زمین بهتر بوده است؛ به‌ویژه از نظر افزایش مقدار R^2 و کاهش نسبی خطاها. در میان مدل‌های بررسی‌شده، CatBoost با $R^2 = 0.182$ ، $Adjusted R^2 = 0.175$ ، $MAE = 0.772$ و همچنین کمترین مقدار RMSE

داده شده است؛ ردیف اول عملکرد مدل را روی داده‌های آزمون و ردیف دوم عملکرد آن را روی داده‌های آموزش نشان می‌دهد. ستون‌های جدول نیز شامل مقادیر چهار معیار ارزیابی -MAE، RMSE، R^2 و Adjusted R^2 - برای هر سناریو و هر مدل هستند. چنین ساختاری امکان تحلیل دقیق و مقایسه‌ای الگوریتم‌ها و سناریوهای داده‌ای را بر اساس عملکرد آن‌ها در داده‌های آزمون فراهم می‌کند. در ادامه، ابتدا نتایج سناریوی داده‌های سطح زمین در بخش ۴-۱، سپس نتایج داده‌های جو بالا در بخش ۴-۲ و سرانجام سناریوی ترکیبی در بخش ۴-۳ ارائه می‌شود.

داده‌های سطح زمین

در این مرحله، پیاده‌سازی مدل‌های یادگیری ماشین بر پایه‌ی داده‌های سطح زمین صورت پذیرفت. پیش از اجرای مدل‌ها، به‌منظور جلوگیری از اثر هم‌خطی میان متغیرها، شاخص VIF برای تمامی ویژگی‌ها محاسبه گردید. نتایج اولیه نشان داد که متغیرهای T_m و U_m به دلیل میانگین‌گیری از مؤلفه‌های دما و رطوبت، با سایر ویژگی‌های گروه خود (T_{min} ، T_{max} و U_{max} و U_{min}) هم‌خطی بالایی دارند و از مجموعه حذف شدند. همچنین در گروه دما، به دلیل هم‌بستگی قوی میان T_{min} و T_{max} ، تنها T_{min} به‌عنوان نماینده‌ی این گروه حفظ گردید؛ زیرا دمای حداقل ارتباط بیشتری با فرایندهای تراکم و وقوع بارش دارد.

پس از حذف این متغیرها، مقادیر VIF سایر ویژگی‌ها در محدوده‌ی قابل قبول قرار گرفت ($VIF < 5$) و هم‌خطی میان متغیرها برطرف شد. در نهایت، مجموعه‌ی نهایی ویژگی‌های به‌کاررفته در مدل‌ها شامل T_{min} ، U_{max} ، U_{min} ، pm ، nm ، ffm و $rrr24$ بوده است. آموزش و ارزیابی مدل‌ها با استفاده از این ویژگی‌ها انجام شد و نتایج عملکرد در جدول ۳ نشان داده شده است.

در داده‌های آزمون، مدل CatBoost با ثبت بالاترین مقادیر R^2 و Adjusted R^2 و کمترین مقدار RMSE، عملکرد برتری نسبت به سایر مدل‌ها نشان داد. در مقابل، مدل AdaBoost با کمترین مقدار MAE بهترین دقت را از منظر خطای مطلق ارائه کرد. دو مدل XGBoost و LightGBM نیز نتایج بسیار نزدیک به یکدیگر و پس از CatBoost به

کاهش چشمگیر مقادیر VIF و ایجاد ساختاری پایدارتر برای داده‌های ترکیبی شد.

پس از اعمال این اصلاحات، تعداد ویژگی‌ها از ۳۱ متغیر اولیه به ۲۲ متغیر کاهش یافت. با این حال، ابعاد مدل همچنان نسبتاً بالا باقی ماند و می‌توانست بر زمان آموزش و قابلیت تفسیرپذیری مدل تأثیر منفی بگذارد. بنابراین، در گام بعدی، به منظور شناسایی مؤثرترین ویژگی‌ها و کاهش بیشتر ابعاد مسئله، از الگوریتم انتخاب ترتیبی پیش‌رو با حذف شناور (SFFS) استفاده شد. این الگوریتم با افزودن و حذف پویای ویژگی‌ها، ترکیبی بهینه از متغیرها را استخراج می‌کند تا تعادل مناسبی میان دقت پیش‌بینی و سادگی مدل برقرار شود. در این پژوهش، معیار انتخاب یا حذف ویژگی‌ها، مقدار R^2 حاصل از اعتبارسنجی متقابل پنج‌تایی در داده‌های آموزش در نظر گرفته شد؛ به طوری که تنها متغیرهایی که بیشترین سهم را در بهبود عملکرد مدل داشتند، در مجموعه‌ی نهایی حفظ شدند و سایر ویژگی‌های کم‌تأثیر حذف گردیدند.

بر پایه نتایج SFFS، منحنی شکل ۱ نشان می‌دهد که با افزایش تعداد ویژگی‌ها، مقدار R^2 به صورت پیوسته تا انتخاب ششمین ویژگی بهبود می‌یابد و پس از آن دچار نوسان یا افت جزئی می‌شود. از این رو، نقطه‌ی ۶ ویژگی به عنوان بهترین تعادل بین دقت و سادگی برگزیده شد؛ انتخابی که ضمن دستیابی به بیشترین R^2 ، از پیچیده‌تر شدن غیرضروری مدل و خطر بیش‌برازش جلوگیری می‌کند. شش ویژگی نهایی عبارت‌اند از $v700$ ، U_{max} ، Spe_Hum500 ، nm و $v500$ و Rel_Hum300 که از میان آن‌ها دو ویژگی U_{max} و nm از داده‌های سطح زمین و چهار ویژگی دیگر از داده‌های جو بالا (ERA5) استخراج گردیده‌اند. حضور هم‌زمان باد نصف‌النهاری در ترازهای ۷۰۰ و ۵۰۰ هکتوپاسکال ($v700$ و $v500$) به نقش انتقال رطوبت با مولفه جنوبی اشاره دارد؛ Spe_Hum500 ظرفیت بخار آب در لایه میانی جو را بازتاب می‌دهد و Rel_Hum300 شرایط مرطوب لایه‌های بالاتر را نشان می‌دهد. در کنار این‌ها، شاخص‌های سطح زمین (nm و U_{max}) اطلاعات تکمیلی درباره‌ی وضعیت سرعت باد نزدیک سطح زمین و ساعت آفتابی را فراهم می‌کنند. این ترکیب، با حداقل پیچیدگی، سیگنال‌های محلی سطحی و متغیرهای ترمودینامیکی در سطوح مختلف جوی را هم‌زمان

$2.294 =$ در داده‌های آزمون، عملکردی پایدار و دقیق از خود نشان داده است. مدل LightGBM نیز رفتار مشابهی داشته و از نظر دقت کلی پس از CatBoost قرار می‌گیرد.

AdaBoost در این سناریو کمترین مقدار $MAE = 0.765$ را ثبت کرده و از این لحاظ بهترین عملکرد را در کاهش خطای مطلق ارائه داده است، هرچند مقدار R^2 آن پایین‌تر از مدل‌های برتر است. در مقابل، جنگل تصادفی با $R^2 = 0.144$ و $RMSE = 2.347$ ضعیف‌ترین دقت را در میان مدل‌ها نشان داده است. به طور کلی، می‌توان نتیجه گرفت که CatBoost در این سناریو به عنوان بهترین مدل شناسایی می‌شود، در حالی که AdaBoost از منظر حداقل خطای مطلق برتر است.

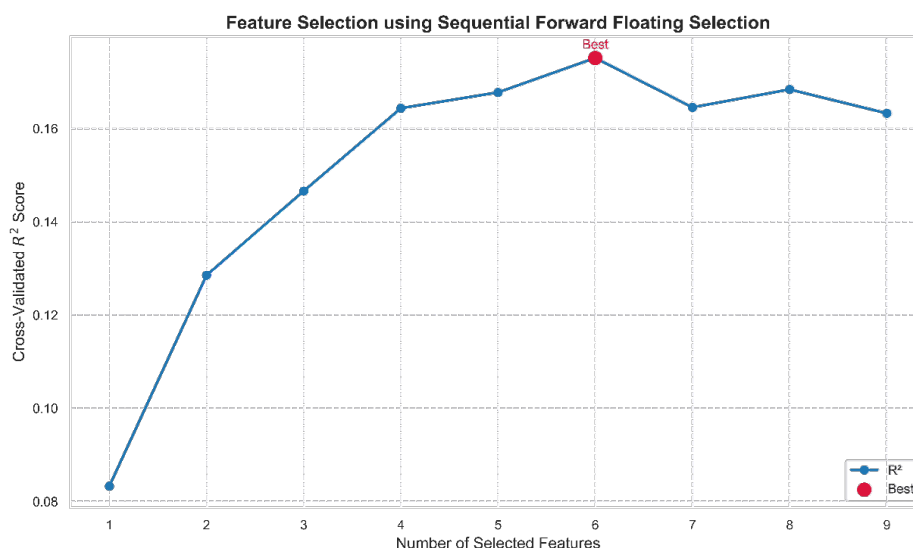
داده‌های ترکیبی

در این سناریو، مدل‌های یادگیری ماشین بر اساس داده‌های ترکیبی حاصل از سطح زمین و لایه‌های بالایی جو توسعه یافتند. در گام نخست، به منظور تضمین پایداری مدل و کاهش اثرات هم‌خطی میان متغیرها، شاخص VIF برای تمامی ویژگی‌ها محاسبه شد. نتایج بررسی اولیه نشان داد که مجموعه‌ای از متغیرها از جمله ارتفاع ژئوپتانسیل (GH) و دما (Temp) در سطوح مختلف جو، همچنین متغیرهای سطحی مانند T_m ، T_{max} ، U_m و pm هم‌خطی بسیار بالایی با سایر ویژگی‌ها دارند. این هم‌بستگی شدید، که ریشه در روابط فیزیکی ذاتی میان متغیرها در سطوح فشاری مختلف و همچنین در نحوه‌ی میانگین‌گیری متغیرهای سطحی دارد، می‌تواند منجر به افزایش واریانس ضرایب و کاهش پایداری مدل گردد.

در مواجهه به این چالش، متغیرهای دارای هم‌خطی بالا به صورت تدریجی از مجموعه‌ی ویژگی‌ها حذف شدند تا افزونگی اطلاعاتی میان آن‌ها کاهش یابد. در این فرآیند، از میان سطوح مختلف ارتفاع ژئوپتانسیل و دما، تنها GH700 و Temp300 به عنوان نمایندگان معنادار حفظ شدند. همچنین، متغیرهای T_m ، T_{max} ، U_m ، pm و Spe_Hum700 به دلیل هم‌پوشانی قابل توجه با سایر پارامترها از مجموعه حذف گردیدند. این پالایش منجر به

است تا اثربخشی هر منبع داده و مزیت‌های بهره‌گیری از رویکرد ترکیبی به صورت دقیق مورد بررسی قرار گیرد.

در مدل لحاظ می‌کند. در نهایت، عملکرد مدل‌های منتخب در سه سناریوی مجزا - شامل داده‌های سطح زمین، داده‌های جو بالا، و رویکرد ترکیبی - با یکدیگر مقایسه و تحلیل شده



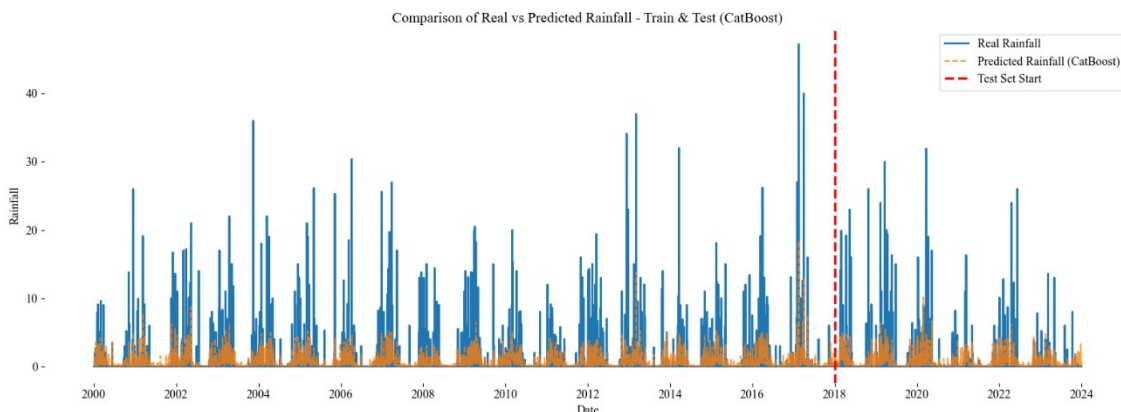
شکل ۱- نتایج حاصل از انتخاب ویژگی‌ها با استفاده از روش SFFS

جو بالا به تصویر کشیده شده‌اند. در شکل‌های مذکور، مقادیر واقعی بارش (میله‌های آبی) در کنار مقادیر پیش‌بینی شده روزانه (میله‌های نارنجی) توسط مدل CatBoost نمایش داده شده است. همچنین، محور افقی تاریخ روز و محور عمودی میزان بارش بر حسب میلی‌متر را نشان می‌دهد و خط چین قرمز آغاز دوره داده‌های آزمون در ۱ ژانویه ۲۰۱۸ را مشخص می‌کند.

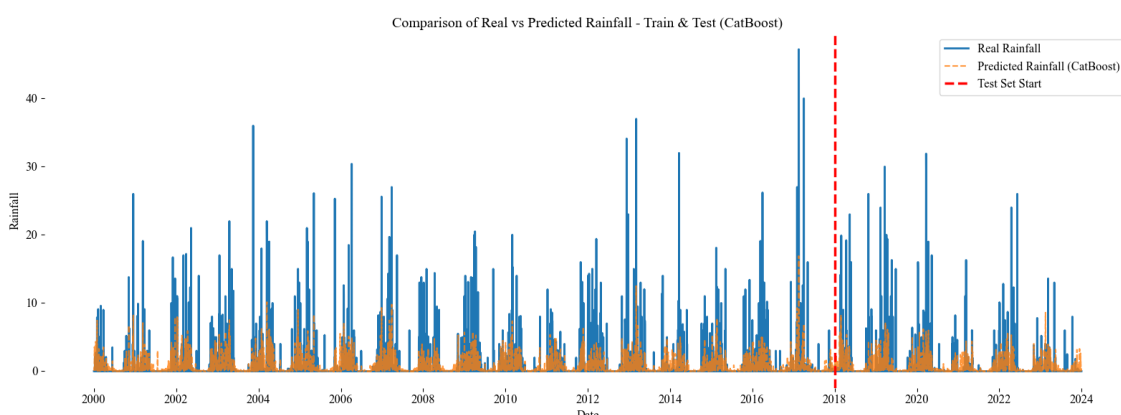
شکل ۲ متعلق به مدل CatBoost است که بر اساس داده‌های سطح زمین آموزش دیده است. این مدل توانسته است الگوی کلی تغییرات بارش را بازتاب دهد و در پیش‌بینی بارش‌های کم و متوسط عملکردی قابل قبول داشته باشد، اما در روزهایی با بارش‌های سنگین، اوج بارش را کمتر از میزان واقعی برآورد کرده است. شکل ۳ نیز نتایج مدل CatBoost را برای داده‌های جو بالا نشان داده است که نسبت به سناریوی سطح زمین دقت بیشتری دارد و نوسانات بارش را بهتر دنبال می‌کند، هرچند همچنان در برآورد دقیق بارش‌های سنگین با خطا روبه‌رو است.

نتایج حاصل از رویکرد ترکیبی "S & U" نشان می‌دهد که این روش با ادغام داده‌های سطح زمین (S) و داده‌های جو بالا (U)، تعادلی مناسب بین دقت و پایداری مدل‌ها ارائه می‌دهد. در میان الگوریتم‌های یادگیری ماشین بررسی شده، CatBoost بهترین عملکرد را در معیارهای Adjusted R²، RMSE و R² به نمایش گذاشته است. به طور خاص، CatBoost در داده‌های آزمون با R² برابر ۰.۱۹۰، Adjusted R² برابر ۰.۱۸۸ و RMSE برابر ۲.۲۸۳ به ترتیب بهترین مقادیر را دارد که نشان‌دهنده توانایی بالای این مدل در تبیین تغییرات داده‌ها و کاهش خطای پیش‌بینی است. در معیار MAE، CatBoost با مقدار ۰.۷۹۵ در داده‌های آزمون پس از AdB Reg. (۰.۷۸۰) رتبه دوم را کسب کرده، که همچنان عملکرد قابل توجهی را نشان می‌دهد. به طور کلی، می‌توان نتیجه گرفت که در این سناریو، الگوریتم CatBoost به عنوان بهترین مدل شناسایی می‌شود، هرچند که AdaBoost از نظر حداقل خطای مطلق عملکرد بهتری داشته است.

در شکل‌های ۲ و ۳، مقادیر واقعی و برآورد شده بارش برای بهترین مدل‌های سناریوهای داده‌های سطح زمین و داده‌های



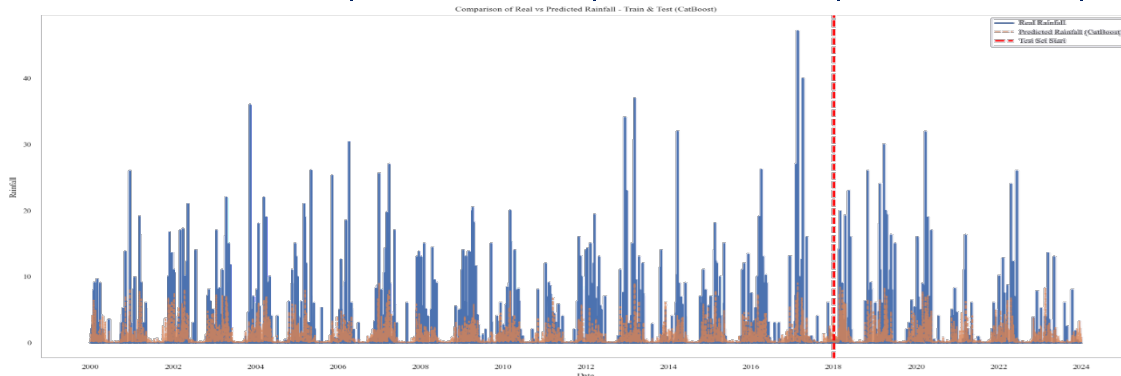
شکل ۲- مقایسه بارش روزانه مشاهده شده و پیش‌بینی شده در ایستگاه مشهد با استفاده از مدل CatBoost بر روی داده‌های S



شکل ۳- مقایسه بارش روزانه مشاهده شده و پیش‌بینی شده در ایستگاه مشهد با استفاده از مدل CatBoost بر روی داده‌های U

جدول ۳- عملکرد پنج مدل یادگیری ماشین در پیش‌بینی بارش روز بعد با استفاده از سه رویکرد S، U و S&U

Methods	MAE			R2			Adjusted R2			RMSE			
	S	U	S & U	S	U	S & U	S	U	S & U	S	U	S & U	
RF Reg.	0.867	0.829	0.818	0.128	0.144	0.174	0.125	0.137	0.172	2.369	2.347	2.305	Test
	0.858	0.823	0.829	0.184	0.310	0.243	0.183	0.308	0.242	2.348	2.160	2.261	Train
AdB Reg.	0.767	0.765	0.780	0.165	0.178	0.178	0.162	0.172	0.176	2.318	2.299	2.300	Test
	0.781	0.776	0.763	0.197	0.327	0.294	0.196	0.325	0.293	2.329	2.132	2.184	Train
XGBoost Reg.	0.855	0.792	0.818	0.162	0.171	0.172	0.159	0.165	0.169	2.322	2.309	2.309	Test
	0.859	0.777	0.776	0.186	0.337	0.338	0.185	0.336	0.338	2.345	2.116	2.114	Train
CatBoost	0.841	0.772	0.795	0.171	0.182	0.190	0.168	0.175	0.188	2.309	2.294	2.283	Test
	0.814	0.765	0.805	0.277	0.331	0.258	0.276	0.329	0.258	2.210	2.126	2.238	Train
LightGBM	0.860	0.790	0.803	0.159	0.181	0.171	0.157	0.175	0.169	2.326	2.295	2.309	Test



شکل ۴- مقایسه بارش روزانه مشاهده شده و پیش‌بینی شده در ایستگاه مشهد با استفاده از مدل CatBoost بر روی داده‌های S&U

شکل ۴ متعلق به مدل CatBoost است که بر اساس داده‌های ترکیبی سطح زمین و جو بالا آموزش دیده است. در این نمودار، مدل توانسته است بیشتر رخدادهای بارش کم تا متوسط را به خوبی پیش‌بینی کند و تغییرات روزانه را تقریباً هم‌زمان با مقادیر واقعی دنبال نماید. با این حال، در زمان‌هایی که بارش بسیار زیاد بوده، پیش‌بینی مدل معمولاً کمتر از مقدار واقعی بوده و به اوج‌ها نرسیده است. دلیل این مسئله می‌تواند نامتوازن بودن داده‌ها باشد، چون تعداد روزهای با بارش کم یا صفر بسیار بیشتر از روزهای با بارش شدید است. با وجود این محدودیت، عملکرد کلی مدل در پیش‌بینی روزانه قابل قبول بوده و می‌تواند برای کاربردهایی که حساسیت زیادی به رخدادهای بسیار شدید ندارند، مفید باشد.

نتیجه‌گیری

در این پژوهش، با استفاده از داده‌های روزانه در بازه زمانی ۲۰۰۰ تا ۲۰۲۳ و در سه سناریوی مختلف (داده‌های سطح زمین، داده‌های جو بالا و ترکیبی از هر دو)، عملکرد چند الگوریتم یادگیری ماشین گروهی برای پیش‌بینی بارش یک روزه (روز بعد) در ایستگاه مشهد مورد ارزیابی قرار گرفت. پیش از مدل‌سازی، به منظور اطمینان از کیفیت و استقلال ویژگی‌های ورودی، اثر هم‌خطی بین متغیرها با استفاده از شاخص VIF بررسی شد و متغیرهایی که مقدار VIF بالاتر از آستانه قابل قبول داشتند، از مجموعه داده‌ها حذف گردیدند. سپس، مدل‌ها تنها با استفاده از ویژگی‌هایی که از نظر آماری مستقل تشخیص داده شده بودند، برازش و ارزیابی شدند. نتایج حاصل از آزمون‌های آماری نشان می‌دهد که سناریوی ترکیبی - که شامل داده‌های هم‌زمان سطح زمین و جو بالا است - در تمامی معیارهای ارزیابی (R^2 , MAE, Adjusted R^2 و RMSE) نسبت به دو سناریوی دیگر عملکرد بهتری داشته است. این یافته حاکی از آن است که تلفیق اطلاعات از لایه‌های مختلف جو، با افزودن عمق فیزیکی و دینامیکی به مدل، می‌تواند دقت پیش‌بینی بارش را به طور قابل توجهی ارتقا دهد.

در میان الگوریتم‌های مورد آزمایش، مدل CatBoost در سناریوی ترکیبی بهترین عملکرد را نشان داد. این سناریو با به‌کارگیری روش انتخاب ویژگی SFFS شکل گرفت و شش

شاخص مؤثر از میان داده‌های اولیه انتخاب شدند که شامل چهار متغیر از سطوح جو بالا ($v700$, U_{max} , Spe_Hum500 و $v500$) و دو متغیر از سطح زمین (nm و Rel_Hum300) بودند. در این چارچوب، مدل CatBoost با کمترین مقدار RMSE معادل ۲.۲۸۳ و هم‌زمان بالاترین مقادیر R^2 برابر ۰.۱۹۰ و R^2 تعدیل‌شده برابر ۰.۱۸۸، برتری خود را در پیش‌بینی بارش روز بعد به وضوح اثبات کرد. این نتیجه گویای توانایی بالای CatBoost در مدل‌سازی روابط غیرخطی و تعاملات پیچیده بین ویژگی‌های چندلایه‌ای جو است، به‌ویژه زمانی که اطلاعات سطح زمین و جو بالا به صورت هدفمند و مبتنی بر انتخاب ویژگی تلفیق شوند.

اگر چه مدل ترکیبی توانست به خوبی افت و خیزهای بارش را پیش‌بینی کند، با این حال، مانند بسیاری از مطالعات مشابه، پیش‌بینی رخدادهای بارش شدید دشوار بود؛ موضوعی که تا حد زیادی به نامتوازن بودن داده‌ها و اتکای ورودی‌ها به اطلاعات «فقط روز قبل» مربوط می‌شود. برای فائق آمدن به این چالش و برای ادامه این خط پژوهشی و بهبود کیفیت آن، می‌توان بعد زمانی ورودی‌ها را با اضافه کردن داده‌های چند روز قبل و همچنین اطلاعات ایستگاه‌های هواشناسی هم‌جوار گسترش داد و شاخص‌هایی ساده مانند میانگین متحرک، بیشینه متحرک، جمع بارش چند روز اخیر و تعداد روزهای پیاپی پربارش یا بی‌بارش را به مدل افزود. همچنین، استفاده از شبکه‌های عصبی دنباله‌ای مانند LSTM و GRU که قادرند اطلاعات چند روز گذشته را هم‌زمان پردازش کنند، می‌تواند وابستگی‌های زمانی بلندمدت‌تری را شناسایی کند. در این حالت، طول بازه زمانی ورودی (مثلاً ۲ تا ۷ روز گذشته) و سایر تنظیمات مدل می‌توانند به‌عنوان پارامترهای قابل تنظیم در فرایند اعتبارسنجی بهینه شوند. علاوه بر این، به‌کارگیری رویکرد دومرحله‌ای - که ابتدا وقوع یا عدم وقوع بارش را پیش‌بینی کرده و سپس تنها در صورت پیش‌بینی بارش، مقدار آن را برآورد کند - می‌تواند دقت پیش‌بینی رخدادهای شدید را افزایش دهد. این رویکرد باعث می‌شود اثر غالب روزهای بدون بارش کاهش یابد و تمرکز مدل روی برآورد شدت بارش بیشتر شود، که در نهایت می‌تواند خروجی مدل را برای سامانه‌های هشدار زودهنگام و مدیریت منابع آب کاربردی‌تر کند. با گسترش این روش پیش‌بینی بر روی همه ایستگاه‌های

- 3448–3461.
<https://doi.org/10.2166/WS.2021.391>
- 9- Fatemeh Mikaili, S. S. (2021). Daily Rainfall Prediction Using Random Forest and Support Vector Regression Methods (Case Study: Ardabil Station). 2118(2021).
<https://doi.org/https://civilica.com/doc/1411181/>
- 10- Freund, Y., & Schapire, R. E. (1999). A Short Introduction to Boosting. In Journal of Japanese Society for Artificial Intelligence (Vol. 14, Issue 5).
- 11- Gaikwad, G. P., & Nikam, P. V. B. (2013). Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model. International Journal of Engineering Research & Technology(IJERT), 2(7), 115–124.
- 12- Hasan, N., Nath, N. C., & Rasel, R. I. (2016). A support vector regression model for forecasting rainfall. 2nd International Conference on Electrical Information and Communication Technologies, EICT 2015, Eict, 554–559.
<https://doi.org/10.1109/EICT.2015.7392014>
- 13- Iliger, S., & Pinto, M. (2022). Comparative Study of Rainfall Prediction Modeling Techniques : Case Study on Karapur, India. In Artificial Intelligence and Communication Technologies (Vol. 7, Issue 3, pp. 159–166). Soft Computing Research Society.
<https://doi.org/10.52458/978-81-955020-5-9-16>
- 14- Khatere Asghari Tahergorabi, Amir Rajabi Behjat, H. D. (2022). Daily Rainfall Prediction using Deep Echo State Network (DeepESN) Method Based on Weather Station Data in Hormozgan Province. 31–38.
- 15- Kisi, O., Heddami, S., Parmar, K. S., Petroselli, A., Külls, C., & Zounemat-Kermani, M. (2025). Integration of Gaussian process regression and K means clustering for enhanced short term rainfall runoff modeling. Scientific Reports, 15(1), 1–26.
<https://doi.org/10.1038/s41598-025-91339-8>
- 16- Li, H., Li, S., & Ghorbani, H. (2024). Data-driven novel deep learning applications for the prediction of rainfall using meteorological data. Frontiers in Environmental Science, 12(August), 1–15.
<https://doi.org/10.3389/fenvs.2024.1445967>
- 17- Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. Journal of Big Data, 8(1), 153.
<https://doi.org/10.1186/s40537-021-00545-4>
- 18- Mahajan, D., & Sharma, S. (2022). Prediction Of Rainfall Using Machine Learning Techniques. 4th International Conference on Emerging Research in Electronics, Computer Science and Technology, ICERECT 2022, 9(01).
<https://doi.org/10.1109/ICERECT56837.2022.10059679>
- 19- Mohammadreza Fallahi, Hadi Varwani, S. G. (2011). Rainfall Prediction Using Tree
- کشور می‌توان پیش‌بینی بارش ۲۴ ساعت آینده کشور را با استفاده از روش‌های یادگیری تهیه کرد. رویکرد ارائه شده در این پژوهش می‌تواند گام اولیه در استفاده از هوش مصنوعی برای پیش‌بینی بارش روزانه کشور بدون استفاده از روش‌های پیش‌بینی عددی وضع هوا باشد.

منابع

- 1- Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7(August 2021), 100204.
<https://doi.org/10.1016/j.mlwa.2021.100204>
- 2- Bian, L., Qin, X., Zhang, C., Guo, P., & Wu, H. (2023). Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM-a case study for runoff simulation in an arid area. Journal of Hydrology, 625(PB), 130091.
<https://doi.org/10.1016/j.jhydrol.2023.130091>
- 3- BREIMAN, L. (2001). Random Forests. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12343 LNCS, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35
- 4- C, S. K., & Sanadi, M. M. (2021). Rainfall Prediction Using Logistic Regression and Support Vector Regression Algorithms. In Communications in Computer and Information Science: Vol. 1440 CCIS. Springer International Publishing. https://doi.org/10.1007/978-3-030-81462-5_54
- 5- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250.
<https://doi.org/10.5194/gmd-7-1247-2014>
- 6- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Aug, 785–794.
<https://doi.org/10.1145/2939672.2939785>
- 7- Dash, Y., Mishra, S. K., & Panigrahi, B. K. (2018). Rainfall prediction for the Kerala state of India using artificial intelligence approaches. Computers and Electrical Engineering, 70(June), 66–73.
<https://doi.org/10.1016/j.compeleceng.2018.06.004>
- 8- Endalieu, D., Haile, G., & Taye, W. (2022). Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia. Water Supply, 22(3),

- 25- Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, 12(2), 1651–1663. <https://doi.org/10.1016/j.asej.2020.09.011>
- 26- Sattari, M. T., Bagheri, R., Shirini, K., & Allahverdipour, P. (2024). Modeling Daily and Monthly Rainfall in Tabriz using Ensemble Learning Models and Decision Tree Regression. *5(18)*, 31–48. <https://doi.org/10.30488/ccr.2024.433394.1192>
- 27- T., A. S. S., Somula, R., K., G., Saxena, A., & A., P. R. (2020). Estimating rainfall using machine learning strategies based on weather radar data. *International Journal of Communication Systems*, 33(13). <https://doi.org/10.1002/dac.3999>
- 28- van der Merwe, J.-P., Wang, T., Clarke, C., & Mansfield, S. D. (2023). Predicting temperature and rainfall for plantation forestry in Mpumalanga, South Africa, using locally developed climate models. *Agricultural and Forest Meteorology*, 329(December 2022), 109275. <https://doi.org/10.1016/j.agrformet.2022.109275>
- Regression Models for Flood Control. <https://doi.org/https://civilica.com/doc/143789/>
- 20- Nguyen, H. N., Nguyen, T. A., Ly, H. B., Tran, V. Q., Nguyen, L. K., Nguyen, M. V., & Ngo, C. T. (2021). Prediction of daily and monthly rainfall using a backpropagation neural Network. *Journal of Applied Science and Engineering*, 24(3), 367–379. [https://doi.org/10.6180/jase.202106_24\(3\).0012](https://doi.org/10.6180/jase.202106_24(3).0012)
- 21- Pande, C. B., Sidek, L. M., Varade, A. M., Elkhachy, I., Radwan, N., Tolche, A. D., & Elbeltagi, A. (2024). Forecasting of meteorological drought using ensemble and machine learning models. *Environmental Sciences Europe*, 36(1). <https://doi.org/10.1186/s12302-024-00975-w>
- 22- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 2018-Decem.
- 23- Qorban Mahtabi, Farshid Taran, S. M. (2018). Prediction of Daily Rainfall Using Meteorological Data from Previous Days (Case Study: Isfahan City).
- 24- Rasol Imani, Reza Ghazavi, A. E. O. (2021). Stochastic Monthly Rainfall Time Series Analysis, Modeling and Forecasting (A case study: Ardebilcity. 84–98.