تحقیقات آب و خاک ایران | دوره ۵۳ | شماره ۲ | اردیبهشت ۱۴۰۱ (ص۳۳۲-۳۱۷)

DOI: https://dx.doi.org/10.22059/ijswr.2022.338036.669197

(مقاله علمی-پژوهشی)

Prediction of Regional Heavy Precipitation Occurrence in the Southwest Iran Using Synoptic Variables and Data Mining Methods

KOKAB SHAHGHOLIAN[1], JAVAD BAZRAFSHAN[1*], PARVIZ IRANNEJAD[2]

1. Department of Irrigation & Reclamation Engineering, Faculty of Agricultural Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Karaj, Tehran, Iran

3. Department of Space Physics, Institute of Geophysics, University of Tehran, Tehran, Iran

(Received: Jan. 26, 2022- Revised: Feb. 11, 2022- Accepted: Feb. 14, 2022)

ABSTRACT

Short-term prediction of heavy precipitation events is especially crucial in flood warning and mitigation. This study offered a novel concept of the regional heavy precipitation based on the probability pattern of a typical rainstorm. Daily precipitation data of 12 synoptic stations located over southwestern Iran were used for this purpose. In addition, six synoptic variables at 1000 to 200 hPa pressure levels on one to five days before heavy precipitations (covering a wide range outside the study area) were used as predictors. All data used in this study cover the period 1987- 2018. Four feature selection methods and 10 binary classifier machine-learning models were employed in this study. The results revealed that using synoptic data up to four days prior to the events best distinguishes heavy precipitation from non-heavy precipitation events. In addition, among the four feature selection methods, Chi-Square and Extra Tree methods are superior to Correlation and Random Forest. As a result of this study, it was found that the Random Forest model with the Chi-Square feature selection method has the highest efficiency in predicting regional heavy precipitation events in the study area. Relative humidity and specific humidity 1-2 days before and wind speed 2-4 days before the precipitation events are relevant synoptic variables for predicting heavy precipitation events.

Keywords: Regional heavy precipitation, Prediction, Data mining, Synoptic variables, Iran.

* Corresponding Author's Email: jbazr@ut.ac.ir

پیش‌بینی رخداد بارش سنگین منطقه‌ای در جنوب غربی ایران با استفاده از متغیرهای همدیدی و روش‌های داده‌کاوی

کوکب شاهقلیان[1]، جواد بذرافشان[1*]، پرویز ایران‌نژاد[2]

۱. گروه آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، تهران، ایران

۲. گروه فیزیک فضا، موسسه ژئوفیزیک دانشگاه تهران، تهران، ایران

(تاریخ دریافت: ۱۴۰۰/۱۱/۶- تاریخ بازنگری: ۱۴۰۰/۱۱/۲۲- تاریخ تصویب: ۱۴۰۰/۱۱/۲۵)

چکیده

پیش‌بینی کوتاه‌مدت بارش‌های سنگین اهمیت ویژه‌ای در هشدار سیل و به‌حداقل‌رساندن آسیب‌های ناشی از آن دارد. در این مطالعه، تعریف جدیدی از بارش سنگین منطقه‌ای برپایه الگوی احتمالاتی رگبارها ارائه شد. برای این منظور از داده‌های بارش روزانه (۲۰۱۸-۱۹۸۷) مربوط به ۱۲ ایستگاه همدید در جنوب غرب ایران استفاده شد. به‌علاوه، شش متغیر همدیدی در ترازهای ۱۰۰۰ تا ۲۰۰ هکتوپاسکال مربوط به یک تا پنج روز قبل از بارش سنگین (که گستره وسیعی در خارج منطقه مطالعاتی را پوشش می‌دهند) به‌عنوان پیش‌بینی گر مورداستفاده قرار گرفت. برای اجرای این پژوهش از چهار روش انتخاب متغیر و ده مدل یادگیری ماشین از نوع طبقه‌بندی‌کننده دودوئی استفاده شد. نتایج نشان داد که به‌منظور تشخیص بارش‌های سنگین از غیر سنگین، بهترین حالت استفاده از داده‌های تا چهار روز پیش از رخداد بارش است. همچنین، از بین چهار روش انتخاب متغیر، روش‌های Chi-Square و Extra Tree بر Correlation و Random Forest برتری دارند. در نتیجه این مطالعه مشخص شد که مدل Random Forest با روش انتخاب متغیر Chi-Square بالاترین کارایی در پیش‌بینی بارش‌های سنگین در منطقه مطالعاتی را دارد. متغیرهای همدیدی مناسب برای پیش‌بینی بارش سنگین شامل رطوبت نسبی و رطوبت ویژه ۱-۲ روز قبل و باد برداری ۲-۴ روز قبل از رخداد بودند.

واژه‌های کلیدی: بارش سنگین منطقه‌ای، پیش‌بینی، داده‌کاوی، متغیرهای همدیدی، ایران.

مقدمه

کشور ایران با مساحتی بالغ بر ۱،۶۴۸،۰۰۰ کیلومترمربع در محدوده جغرافیایی ۴۴ تا ۶۴ درجه شرقی طول و ۲۵ تا ۴۰ درجه شمالی عرض جغرافیایی واقع شده است. اگرچه این کشور از جنوب با خلیج فارس و دریای عمان و از شمال با دریای خزر هم مرز است اما در یکی از کمربندهای خشک زمین قرار دارد. بیابان‌های پرفشار نیمه گرمسیری و داخلی باعث ایجاد شرایط خشک و نیمه خشک برای حدود ۷۵ درصد از مساحت کشور شده است (Rahimi and Fatemi, 2019). بارش‌های سنگین سهم کمی از تعداد روزهای بارشی ایران را شامل می‌شوند. با این حال، این رویدادها منبع اصلی تأمین آب کشور هستند (Alijani et al., 2008).

بارش‌های سنگین از جمله رخدادهای فرین هوا-آب‌شناسی به شمار می‌روند. طبق تعریف سازمان جهانی هواشناسی (WMO[1])، یک رخداد بارندگی مشخص که در طی یک دوره زمانی ۱، ۳، ۶، ۱۲، ۲۴ و یا ۴۸ ساعته اتفاق می‌افتد و میزان کل بارندگی حاصل از آن برای یک مکان معین (بر مبنای ایستگاه) از آستانه معینی (صدک ۹۰ ام، صدک ۹۵ ام، یا صدک ۹۹ ام) بیشتر باشد، بارش سنگین نامیده می‌شود. سازمان جهانی هواشناسی پیشنهاد داده است که به‌عنوان یک استاندارد مشترک، از مقیاس زمانی ۲۴ ساعته استفاده شود (WMO, 2016).

طبق گزارش چهارم مجمع بین‌دولتی تغییر اقلیم (IPCC[2])، امروزه شواهد علمی زیادی مبنی بر این که گرمایش زمین باعث افزایش فراوانی رخدادهای فرین اقلیمی و هوا-آب‌شناسی مانند خشک‌سالی و بارش سنگین در آینده خواهد شد، وجود دارد (IPCC, 2007). ایران نیز از این شرایط مستثنی نبوده و در پژوهش‌های انجام شده به این مساله اشاره شده است. به‌عنوان نمونه Rahimi و Fatemi (2019) بارش‌های میانگین و فرین ۳۳ ایستگاه همدید در سراسر ایران را طی دوره ۵۸ ساله ۲۰۱۷-۱۹۶۰ بررسی کردند. آنها روند افزایشی قابل توجهی در مقدار،

* نویسنده مسئول: jbazr@ut.ac.ir

1 World Meteorological Organization
2 Intergovernmental Panel on Climate Change

شدت و فراوانی بارندگی‌های فرین، به ویژه در مناطق جنوب غربی ایران و مناطق ساحلی خلیج فارس یافتند. بارش شدید باران می‌تواند منجر به جاری شدن ناگهانی رواناب و سیل فاجعه بار شود.

اگرچه در دو دهه گذشته برای به‌حداقل‌رساندن اثرات، خطرات و تلفات رخدادهای فرین، شاهد پیشرفت‌های تحقیقاتی قابل توجهی در دنیا بوده‌ایم (Cavazos et al., 2008; IPCC, 2012; Wheater, 2002; Young, 2002)، اما همچنان تعداد قابل توجهی از این وقایع، موجب تلفات عظیم انسانی و اقتصادی می‌شود. دفتر هماهنگی امور بشردوستانه سازمان ملل متحد (OCHA[1]) در ایران، در گزارش ۱۰ مارس ۲۰۲۰، اعلام کرد که بارندگی‌های شدید از ۲۴ فوریه تاکنون باعث سیل گسترده در جنوب غرب ایران از جمله استان‌های لرستان و خوزستان شده است[2]. آسیب به چندین جاده و تخریب پل دسترسی به حداقل ۵۸ روستا از جمله عواقب این سیل به شمار می‌رود. همچنین بیش از ۱۵۰ روستا بر اثر سیلاب شدید در لرستان و اطراف آن دچار قطعی گاز شده و تمامی راه‌ها مسدود شده است. بارش شدید باران از ۵ تا ۷ دسامبر سال ۲۰۲۰ باعث جاری شدن سیل در ایران شد[3]. بارش شدید باران تا ۱۶۰ میلی‌متر طی چند ساعت در دشتستان در استان بوشهر گزارش شد. همزمان سیل و بارندگی شدید در شهرستان آباده در استان فارس نیز در این بازه زمانی گزارش شده است. به‌عنوان نمونه‌ای دیگر، شهرستان آغاجاری در استان خوزستان در ۲۴ تا ۲۹ دسامبر ۲۰۲۰، ۹۰ میلی‌متر بارندگی را به ثبت رسانده است. این در حالی است که شهرهای ایلام در استان ایلام و صفی‌آباد در استان خوزستان هر دو ۶۸ میلی‌متر بارندگی را در همین مدت ثبت کرده‌اند[4]. در طی این بارش‌های فرین، بیش از ۵۰ خانه آسیب دیده و ۲۶۳ نفر تخلیه شده و اسکان موقت داده شده‌اند. همچنین سیل‌های رخ داده در ۴ فوریه ۲۰۰۶ در استان لرستان (Arvin and Mohamadinejad, 2015)، ۲۰ نوامبر ۲۰۱۱ در استان‌های خوزستان و کهکیلویه و بویراحمد (Khoshakhlagh et al., 2015) و ۱۵ تا ۱۷ فوریه سال ۲۰۱۷ در استان بوشهر (Vaghefi et al., 2019) از نمونه‌های دیگر به شمار می‌روند. تلفات جانی، وارد آمدن خسارات زیاد بر تأسیسات زیربنایی و لوله‌های انتقال نفت، تخریب واحدهای مسکونی و زمین‌های کشاورزی، قطع شبکه‌های آبرسانی، برق و مخابرات، قطع راه‌های ارتباطی و محاصره شدن روستاها در سیلاب، از جمله اثرات مخرب این سیلاب‌ها بوده است.

به‌منظور به‌حداقل‌رساندن آسیب‌های ناشی از بارش‌های فرین و سیل‌های به وقوع پیوسته ناشی از این بارش‌ها، هشدار سیل از اهمیت ویژه‌ای برخوردار است. همچنین درک تغییرات در بزرگی و فرکانس بارش شدید برای ارزیابی خطر سیل آینده و مدیریت منابع آب ضروری است (Beguería et al., 2011; Seneviratne et al., 2012; Mallakpour and Villarini, 2015; Hirsch and Archfield, 2015).

یکی از مؤلفه‌های اصلی برای هشدار سیل، پیش‌بینی کوتاه‌مدت رخداد بارش سنگین است که حتی با بهبود مدل‌های عددی پیش‌بینی وضع هوا، همچنان به‌عنوان یک چالش باقی مانده است. این نوع پیش‌بینی یک چالش تحقیقاتی با اولویت بسیار بالا به خصوص برای مناطق دارای سکنه و مستعد سیل می‌باشد (Nayak and Ghosh, 2013). باتوجه‌به پیچیدگی سامانه اقلیم و نگرانی‌های اجتماعی در مورد اثرات تغییرات آب و هوایی، نیاز به توسعه و استفاده از روش‌های پیشرفته و دقیق‌تر در این زمینه احساس می‌شود. امروزه داده‌کاوی[5] به‌عنوان یکی از فنون پیشرفته و دقیق‌تر در مقایسه با روش‌های پیشین پیش‌بینی شناخته شده است. هدف اصلی داده‌کاوی این است که از مجموعه داده‌های در دسترس، اطلاعات استخراج کرده و آن‌ها را به یک ساختار قابل درک به‌منظور تسهیل تفسیر داده‌های موجود تبدیل کند (Fayyad et al., 1996). از این تکنیک‌ها می‌توان برای استخراج دانش پنهان از داده‌های سری زمانی برای استفاده در آینده استفاده کرد (Ahmad et al., 2017a; Mishra et al., 2017; Aftab et al., 2018; Gupta et al., 2018). فنون داده‌کاوی ظرفیت استخراج الگوهای پنهان در داده‌های آب و هوای گذشته را دارند و می‌توانند با استفاده از الگوهای استخراج شده، شرایط آب و هوایی آینده را پیش‌بینی کنند (Aftab et al., 2018). چندین دهه است که اقلیم‌شناسان از فنون داده‌کاوی در مطالعات مختلف استفاده می‌کنند. با این حال در چارچوب خاص مرتبط با وقایع بارندگی فرین، فنون داده‌کاوی در تعداد نسبتاً کمی از مطالعات استفاده شده است (Ruivo et al., 2015). به طور کلی، روش‌های داده‌کاوی رویکردی امیدوارکننده در بررسی رویدادهای فرین هوا- آب شناسی و استخراج دانش از مجموعه بزرگ و پیچیده داده‌ها به شمار می‌رود (Ruivo et al., 2015). در یک بررسی جامع روی معماری‌های مختلف شبکه عصبی که برای پیش‌بینی بارندگی در ۲۵ سال گذشته استفاده شده است، نویسندگان تاکید کردند که اکثر محققان با استفاده از تکنیک‌های

1 United Nations Office for the Coordination of Humanitarian Affairs
2 https://floodlist.com/asia/iran-floods-lorestan-khuzestan-march-2020#
3 https://floodlist.com/asia/iran-floods-december-2020
4 https://floodlist.com/asia/iran-floods-november-2020
5 Data mining

پیش‌بینی مانند SVM[1]، MLP[2]، BPN[3]، RBFN[4] و SOM[5] نتایج قابل توجهی در پیش‌بینی بارندگی بدست آورده‌اند و این تکنیک‌ها مناسب‌تر از سایر تکنیک‌های آماری و عددی هستند (Nayak et al., 2013).

Abbot و Marohasy (2014) تعدادی از مدل‌های پیش‌بینی یادگیری ماشین (ML[6]) با مدل‌های پیش‌بینی فیزیکی را برای پیش‌بینی بارش ماهانه مقایسه کردند و نشان دادند که دقت مدل‌های پیش‌بینی یادگیری ماشین بالاتر است. در یک تحلیل مقایسه‌ای از ماشین بردار پشتیبان (SVM)، شبکه‌های عصبی مصنوعی (ANN[7]) و سیستم استنتاج عصبی فازی تطبیقی (ANFIS[8]) در زمینه پیش‌بینی بارش، مشخص شد که هنگامی که مدل‌ها فقط با داده‌های بارش فرین آموزش داده می‌شوند، مدل ANN بهترین نتایج را ارائه می‌دهد، اما برای پیش‌بینی طوفان‌های شدید مدل SVM توصیه می‌شود (Zhang et al., 2016). به‌منظور پیش‌بینی بارش در مالزی، مقایسه‌ای بین فنون مختلف داده‌کاوی مانند جنگل تصادفی (RF[9])، SVM، بیز ساده (NB[10])، ANN و درخت تصمیم (DT[11]) انجام شد. نتایج، عملکرد قابل توجه جنگل تصادفی را نشان داد زیرا این تکنیک مقادیر زیادی از نمونه‌ها را با داده‌های آموزشی کم به درستی طبقه‌بندی کرد (Zainudin et al., 2016). کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه همدید کرمانشاه را در دوره آماری ۱۳۴۹ تا ۱۳۸۹ ارزیابی شد. نتایج نشان داد که مدل درخت تصمیم رگرسیونی، مدلی به‌نسبت کارا در پیش‌بینی بارش می‌باشد که استفاده از میانگین متحرک منجر به افزایش چشمگیر کارایی این مدل می‌شود (Omidvar et al., 2014). با استفاده از مدل درخت تصمیم در پیش‌بینی بارش ایستگاه ساری این نتیجه بدست آمد که درخت تصمیم CART یک روش مناسب برای پیش‌بینی‌های بلندمدت هواشناسی با استفاده از داده‌های گذشته است (Baharian and Salimi, 2018). Poursalehi et al. (2019) پژوهشی انجام دادند که هدف آن پیش‌بینی بارش ماهانه با به‌کارگیری الگوریتم‌های داده‌کاوی درخت تصمیم (M5) و K-نزدیک‌ترین همسایه (KNN[12]) و مقایسه این دو روش در راستای تعیین روش کارامدتر در زمینه پیش‌بینی بارش بود. نتایج نشان داد که در تمامی سناریوهای تعریف شده، مدل درختی M5 نسبت به مدل KNN توانایی بیشتری در پیش‌بینی بارش ماهانه این ایستگاه دارد.

با توجه به آن‌چه گفته شد می‌توان به اهمیت پیش‌بینی بارش‌های سنگین با استفاده از روش‌های داده‌کاوی به منظور افزایش دقت پیش‌بینی پی برد. هدف پژوهش حاضر مشخص‌کردن بهترین ترکیب متغیرهای همدیدی و یافتن مدل یا مدل‌های داده‌کاوی مناسب برای پیش‌بینی دقیق‌تر بارش‌های سنگین در منطقه جنوب غرب ایران است.

داده‌ها و روش‌ها

داده‌های بارش

داده‌های بارش روزانه ۱۲ ایستگاه همدیدی از بایگانی داده‌های سازمان هواشناسی کشور برای سال‌های ۱۹۸۷ تا ۲۰۱۸ دریافت گردید. ایستگاه‌های موردمطالعه طوری انتخاب شده‌اند که پراکندگی خوبی در منطقه موردمطالعه (جنوب غرب ایران) داشته و دارای کمترین میزان داده گم‌شده باشند. اگرچه حدود ۹۰ درصد بارش سالانه ایران از مهر تا اردیبهشت (به استثنای سواحل دریای خزر که ۷۵ درصد است) رخ می‌دهد (Pourasghar et al., 2021)، برخی بارش های سنگین خارج از دوره مهر تا اردیبهشت رخ داده است. ایستگاه‌های آبادان و شهرکرد با ۶/۶ و ۲۰۴۹ متر ارتفاع از سطح دریا به ترتیب پست‌ترین و مرتفع‌ترین ایستگاه‌ها هستند. ایستگاه ایلام با ۲۸۲ میلی‌متر بارش در ۲۹ اکتبر ۲۰۱۵ بیشترین بارش ۲۴ ساعته (روزانه) را در طی دوره آماری در بین ۱۲ ایستگاه دریافت کرده است. کمترین بارش حداکثر ۲۴ ساعته در طی این دوره مربوط به ایستگاه آبادان برابر با ۵۴ میلی‌متر است. مشخصات و پراکندگی جغرافیایی ایستگاه‌های همدیدی مورداستفاده در این پژوهش به ترتیب در جدول ۱ و شکل ۱ آورده شده است.

متغیرهای همدیدی (پیش‌بینی گرها)

با بهره‌گیری از تارنمای سازمان ملی اقیانوسی و جوی آمریکا[13](NOAA[14])، انحراف از میانگین روزانه متغیرهای همدیدی با تفکیک افقی ۲/۵ × ۲/۵ درجه طول و عرض جغرافیایی با استفاده از داده‌های بازتحلیل مراکز ملی پیش‌بینی محیطی (NCEP[15]) و مرکز ملی پژوهش‌های جوی (NCAR[16])، برای پهنه موردمطالعه تهیه شدند. مشخصات متغیرهای همدیدی

1 Support Vector Machine
2 Multi Layer Perceptron
3 Back Propagation Network
4 Radial Basis Function Networks
5 Self Organizing Map
6 Machin Learning
7 Artificial Neural Networks
8 Adaptive Neuro Fuzzy Inference System
9 Random Forest
10 Naive Bayes
11 Decision Tree
12 K- Nearest Neighbors
13 http://www.esrl.noaa.gov/psd/data/composites/day/
14 National Oceanic and Atmospheric Administration
15 National Centers for Environmental Prediction
16 National Center for Atmospheric Research

موردمطالعه در جدول ۲ آورده شده است. همان طور که در جدول مشاهده می‌شود، چهار متغیر همدید دمای هوا (Air temperature)، ارتفاع ژئوپتانسیل (Geopotential height)، اُمگا (Omega) و باد برداری (Vector wind) در پنج تراز فشاری ۱۰۰۰، ۸۵۰، ۷۰۰، ۵۰۰ و ۲۰۰ میلی‌بار و دو متغیر همدید رطوبت ویژه (Specific Humidity) و رطوبت نسبی (Relative Humidity) در ترازهای فشاری نام برده به غیر از ۲۰۰ میلی‌بار، در این پژوهش به‌عنوان متغیرهای مستقل مورد استفاده قرار گرفتند. این متغیرها کل محدوده جغرافیایی مشخص شده در شکل ۱ را پوشش می‌دهند.

جدول ۱- مشخصات ایستگاه‌های همدیدی مورد استفاده

شماره ایستگاه	نام ایستگاه	عرض جغرافیایی	طول جغرافیایی	ارتفاع از سطح دریا (متر)	اقلیم (دومارتن)*	تاریخ حداکثر بارش ۲۴ ساعته	حداکثر بارش ۲۴ ساعته (میلی‌متر)
۱	آبادان	۳۰° ۲۲´	۴۸° ۱۵´	۶/۶	بسیار خشک	۲۰۰۵/۱۲/۱۷	۵۴
۲	آباده	۳۱° ۱۱´	۵۲° ۴۰´	۲۰۳۰	خشک	۲۰۰۳/۱۲/۰۶	۸۳
۳	اهواز	۳۱° ۲۰´	۴۸° ۴۰´	۲/۵۲	خشک	۱۹۹۷/۱۱/۱۱	۱۰۷
۴	بندرعباس	۲۷° ۱۳´	۵۶° ۲۲´	۹/۸	بسیار خشک	۲۰۰۰/۰۱/۱۷	۱۲۸
۵	بوشهر	۲۸° ۵۹´	۵۰° ۵۰´	۱۹/۶	خشک	۲۰۰۲/۰۱/۱۱	۱۴۴
۶	فسا	۲۸° ۵۸´	۵۳° ۴۱´	۱۲/۳۸۸	نسبتاً خشک	۱۹۹۵/۰۱/۰۶	۱۱۰
۷	ایلام	۳۳° ۳۸´	۴۶° ۲۶´	۱۳۳۷	نسبتاً خشک	۲۰۱۵/۱۰/۲۹	۲۸۲
۸	خرم‌آباد	۳۳° ۲۶´	۴۸° ۱۷´	۱۱/۸۴۷	نسبتاً خشک	۲۰۱۶/۰۴/۱۳	۹۴
۹	مسجدسلیمان	۳۱° ۵۶´	۴۹° ۱۷´	۳/۵۲۰	نسبتاً خشک	۱۹۹۳/۰۲/۲۵	۱۵۷
۱۰	شهرکرد	۳۲ °۱۷´	۵۰° ۵۱´	۲۰/۹۴۸	مرطوب	۲۰۰۶/۱۱/۱۳	۸۸/۸
۱۱	شیراز	۲۹° ۳۲´	۵۲° ۳۶´	۱۴۸۴	نسبتاً مرطوب	۱۹۹۲/۱۲/۲۱	۷۵
۱۲	یاسوج	۳۰° ۵۰´	۵۱° ۴۱´	۱/۵۸۳۱	مدیترانه‌ای	۲۰۰۴/۰۱/۱۱	۱۳۵

* (Khalili and Rahimi, 2018)

جدول ۲- مشخصات متغیرهای همدید مورد مطالعه

Variable	symbol	Mean/Anomaly	Unit	Pressure levels (hPa)
Air temperature	air	Anomaly	K	200, 500, 700, 850 and 1000
Geopotential height	hgt	Anomaly	m	200, 500, 700, 850 and 1000
Omega	omega	Anomaly	Pa/s	200, 500, 700, 850 and 1000
Relative Humidity	rhum	Anomaly	--	500, 700, 850 and 1000
Specific Humidity	shum	Anomaly	g/kg	500, 700, 850 and 1000
Vector wind	uwnd & vwnd	Anomaly	m/s	200, 500, 700, 850 and 1000

روش‌های پردازش

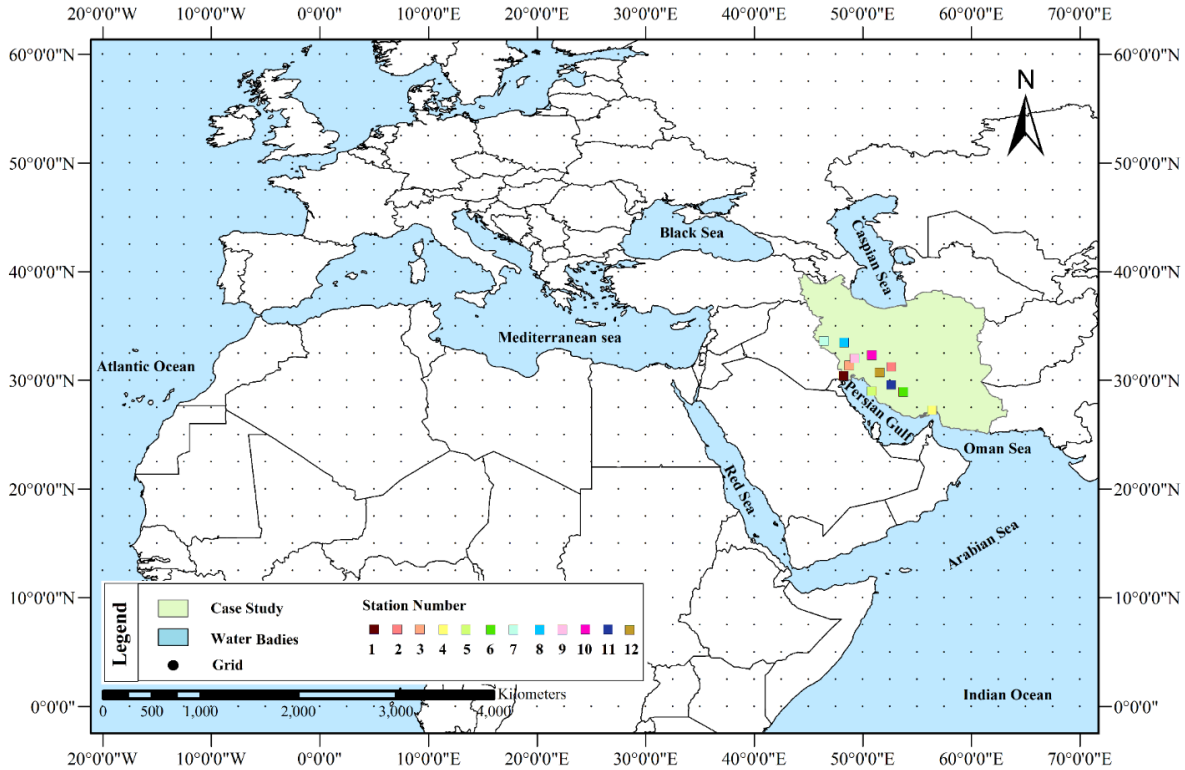
انتخاب روزهای بارش سنگین در منطقه مطالعاتی

در بیشتر مطالعات انجام شده برای مشخص‌کردن روزهای دارای بارش سنگین از فراسنج صدک‌ها و به صورت ایستگاهی استفاده شده است (Groisman et al. 2005; Nazemosadat and Shahgholian 2017; Sun et al. 2020). به‌طور معمول، در تعریف بارش سنگین از آستانه‌های صدکی ۹۹، ۹۵ یا ۹۰ استفاده می‌شود. در این پژوهش از روشی ابتکاری برای انتخاب روزهای

بارش سنگین منطقه‌ای استفاده شد. اگر یک الگوی مکانی بارش سنگین (استورم) را در نظر بگیریم، بیشترین بارش در مرکز آن اتفاق می‌افتد. مرکز الگوی مکانی بارش سنگین می‌تواند در هر بخشی از منطقه مطالعاتی قرار بگیرد. به‌تدریج که از مرکز الگوی مکانی بارش سنگین دورتر می‌شویم از مقدار بارش کاسته می‌شود. در این مطالعه، با استفاده از هر سه آستانه صدکی مذکور، روزهای بارش سنگین منطقه‌ای روزهای هستند که بارش دست کم در نیمی از ایستگاه‌ها باریده و مقدار آن در دست‌کم یک ایستگاه بزرگتر از صدک نود و نهم، در دست‌کم دو ایستگاه بزرگتر از صدک نود و پنجم و در دست‌کم سه ایستگاه بزرگتر از صدک نودم است. بدین ترتیب، ۹۲ روز به‌عنوان روزهای بارش سنگین در منطقه مطالعاتی در طی دوره آماری ۲۰۱۸-۱۹۸۷ تشخیص داده شد.

انتخاب روزهای بارش غیر سنگین

به‌منظور انتخاب روزهای بارش غیر سنگین، از آنجایی که ممکن است تا چند روز قبل و بعد از بارش‌های سنگین، بارش‌های غیر سنگینی توسط همان سیستم همدیدی روی منطقه داشته باشیم؛ علاوه بر روزهای بارش سنگین، سه روز قبل و سه روز بعد از بارش‌های سنگین (چه دارای بارش باشند چه نباشند) نیز از روزها حذف و از باقی‌مانده روزهای دارای بارش به صورت تصادفی، روزهای بارش غیر سنگین را انتخاب می‌کنیم. از آنجایی که اگر تعداد روزهای بارش سنگین و غیر سنگین با هم در تعادل نباشد و تعداد روزها در یکی از آنها بیشتر از دیگری باشد، روش‌های داده‌کاوی بیشتر بر طبقه‌بندی نمونه بزرگ‌تر تمرکز کرده و نمونه کوچک‌تر را نادیده گرفته یا به اشتباه طبقه‌بندی می‌کنند (Longadge et al., 2013)؛ بدین ترتیب ۹۲ روز تصادفی به‌عنوان روزهای بارش غیر سنگین انتخاب شدند.



شکل ۱- موقعیت ایران نسبت به منابع آب موجود در محدوده مورد مطالعه و پراکندگی جغرافیایی ایستگاه‌های همدیدی در جنوب غرب ایران

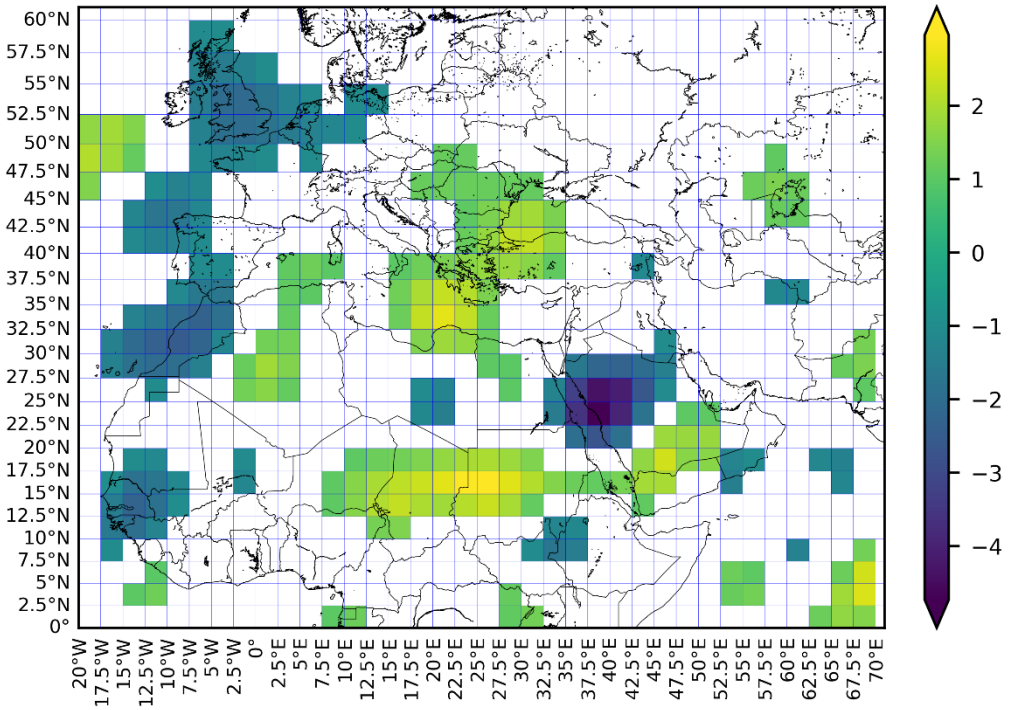
تعیین گریدهای بی‌هنجاری شدید در محدوده جغرافیایی

به‌منظور پایش دقیق بدنه‌های آبی موثر بر بارش‌های سنگین در ایران، پا را فراتر از خاور میانه گذاشته و محدوده موردمطالعه را طوری انتخاب می‌کنیم که علاوه بر دریای عمان، خلیج فارس، دریای عرب، دریای سرخ و دریای مدیترانه، بخش‌هایی از اقیانوس‌های اطلس و هند را نیز در بر بگیرد (شکل ۱). محدوده موردمطالعه، موقعیت جغرافیایی ایران و آب‌های اطراف آن که منبع رطوبت برای بارش‌های کشور محسوب می‌شوند و همچنین موقعیت جغرافیایی ایستگاه‌های همدید در جنوب غرب ایران در این شکل قابل مشاهده است. گریدهای بی هنجاری شدید یعنی مجموعه گریدهای منطقه مطالعاتی (محدوده جغرافیایی ۰ تا ۶۰ درجه شمالی و ۲۰ درجه غرب تا ۷۰ درجه شرق) که تفاوت معنی‌داری را از نظر مقادیر متغیرهای همدیدی مختلف در هنگام

وقوع بارش سنگین و بارش غیر سنگین نشان می‌دهند. بدین منظور، ابتدا نقشه‌های همدیدی که نشان دهنده اختلاف مقادیر هر یک از متغیرها بین روزهای بارش سنگین و بارش غیر سنگین هستند، در ترازهای مختلف تهیه شدند. مقادیر هر متغیر در هر یک از این نقشه‌ها به مقادیر استاندارد تبدیل شد. گریدهایی که یک انحراف معیار بزرگ‌تر (کوچک‌تر) از میانگین هستند به‌عنوان گریدهای بی‌هنجاری شدید مثبت (منفی) معرفی شدند. این روش برای همه متغیرها در همه ترازهای فشاری و برای یک تا پنج روز پیش از رخداد بارش‌های سنگین انجام شد. گریدهای بی هنجاری شدید مشخص شده برای هر متغیر، متفاوت از متغیر دیگر و در یک متغیر مشخص نیز در ترازهای مختلف فشاری متفاوت هستند. به‌عنوان نمونه آن‌چه که در شکل ۲ مشاهده می‌شود گریدهای بی‌هنجاری شدید متغیر امگا برای دو روز پیش از رخداد بارش سنگین در تراز ۲۰۰ میلی‌بار است. برای استفاده عملیاتی از گریدهای بی‌هنجاری شدید در پیش‌بینی بارش سنگین جنوب غرب ایران، میانگین گریدها در سه حالت (کل گریدها، کل گریدهای بی‌هنجاری منفی و کل گریدهای بی‌هنجاری مثبت) محاسبه شد. بنابراین، برای هر متغیر همدیدی در هر تراز فشاری از یک تا ۵ روز قبل از وقوع بارش سنگین/غیر سنگین، سه نوع میانگین در هر روز محاسبه شد. به تبع این سه نوع روش میانگین‌گیری، فرایند مدل‌سازی رابطه بین بارش‌های سنگین/غیر سنگین و متغیرهای همدیدی نیز سه بار تکرار شد: ، یکبار با میانگین گریدهای بی‌هنجاری شدید منفی (حالت Negative)، بار دوم با میانگین گریدهای بی‌هنجاری شدید مثبت (حالت Positive)، و بار سوم با میانگین کل گریدهای بی‌هنجاری شدید مثبت و منفی (حالت Both).

انتخاب متغیرها

انتخاب متغیر، فرآیند کاهش تعداد متغیرهای ورودی به مدل پیش‌بینی با هدف حذف هم‌خطی بین متغیرهای پیش‌بینی گر است. به بیان دیگر، روش‌های انتخاب متغیر، پیش‌بینی‌گرهای ورودی را غربال نموده و به تعدادی که مطلوب مدل پیش‌بینی متغیر وابسته (در اینجا رخداد/عدم رخداد بارش سنگین منطقه‌ای مد نظر می‌باشد) است، کاهش می‌دهند. کاهش تعداد متغیرهای ورودی برای کاهش هزینه محاسباتی مدل‌سازی و در بعضی موارد برای بهبود عملکرد مدل، مطلوب است (Hall, 2000). در این پژوهش، به‌منظور شناسایی تاثیرگذارترین متغیرهای همدیدی در طبقه‌بندی بارش‌های سنگین و غیر سنگین از چهار روش Correlation (روش A)، Chi-Square (روش B)، Random Forest (روش C) و Extra Tree (روش D) استفاده شد که چهار روش مورد نظر با حالت عدم استفاده از روش‌های انتخاب متغیر (Without Feature Selection) (روش E) مقایسه شدند. روش‌های انتخاب متغیر و توضیح مختصری درباره آن‌ها در جدول ۳ آورده شده است.



شکل ۲- گریدهای بی‌هنجاری شدید متغیر امگا برای دو روز پیش از رخداد بارش سنگین در تراز ۲۰۰ میلی‌بار

مدل‌های داده‌کاوی مورداستفاده

الگوریتم‌های داده‌کاوی به دودسته نظارت شده و بدون نظارت طبقه‌بندی می‌شوند. روش‌های نظارت شده ابتدا با استفاده از بخشی از داده‌های ورودی آموزش می‌بینند و سپس با بخش دیگری از داده‌های ورودی ارزیابی می‌شوند (Ahmad and Aftab 2017; Ahmad et al. 2017a; Ahmad et al. 2017b).

از آنجایی که مساله پیش روی ما یک مشکل طبقه‌بندی است و فقط دو گزینه بارش سنگین و غیر سنگین را داریم، از مدل‌هایی استفاده می‌کنیم که مدل‌های طبقه‌بندی نامیده شده و پیش‌بینی دو کلاس را پشتیبانی می‌کنند. الگوریتم‌هایی که در این پژوهش مورداستفاده قرار گرفتند و توضیح مختصری در مورد هرکدام به طور جداگانه در جدول ۴ مشاهده می‌شوند.

در فرآیند مدل‌سازی بارش سنگین/غیر سنگین، مدل‌های ارائه‌شده در جدول ۴ با هر چهار روش انتخاب متغیر (جدول ۳) اجرا شد. در اینجا ۷۵ درصد از داده‌ها برای واسنجی مدل‌ها و ۲۵ درصد مابقی برای اعتبارسنجی استفاده می‌شود. به این ترتیب از مجموع ۱۸۴ روز بارش سنگین و غیر سنگین، ۱۳۸ روز برای آموزش مدل‌ها و ۴۶ روز نیز برای آزمون آن‌ها مورداستفاده قرار می‌گیرد. سناریوی تعریف شده به این ترتیب است که مدل ابتدا برای یک روز پیش از رخداد بارش سنگین، سپس برای یک و دو روز پیش از رخداد بارش سنگین، و به همین ترتیب تا ۵ روز پیش از وقوع بارش سنگین اجرا می‌شود. هدف از اجرای این سناریو یافتن پاسخ برای این پرسش است که آیا با اضافه کردن روزهای دوم، سوم، چهارم و پنجم پیش از رخداد بارش به یک روز پیش از رخداد، دقت پیش‌بینی مدل‌ها افزایش می‌یابد؟ همچنین با اجرای این سناریو میزان اهمیت n-امین (n از ۱ تا۵) روز پیش از رخداد بارش، در طبقه‌بندی بارش‌های سنگین و غیر سنگین مشخص می‌شود.

جدول ۳- روش‌های انتخاب متغیر و توضیح مختصری درباره آن‌ها

روش انتخاب متغیر	توضیح
Correlation (A)	همبستگی معیاری از رابطه خطی دو یا چند متغیر است (مثلاً x و y که به‌عنوان $x = 2y$ به یکدیگر بستگی دارند). منطق استفاده از همبستگی برای انتخاب ویژگی این است که متغیرهای خوب با هدف ارتباط زیادی دارند. علاوه بر این‌که متغیرها باید با هدف در ارتباط باشند باید بین خودشان بی‌ارتباط باشند. اگر دو متغیر با هم ارتباط داشته باشند، تقریباً تأثیر یکسانی بر متغیر وابسته دارند. بنابراین، وقتی دو ویژگی همبستگی بالایی دارند، می‌توانیم یکی از این دو ویژگی را کنار بگذاریم (Hall, 2000).
Chi-Square (B)	این روش انتخاب ویژگی، در زیرمجموعه روش‌های انتخاب ویژگی نظارت شده قرار می‌گیرد. بین هر ویژگی و هدف، Chi-square محاسبه شده و ویژگی‌های با بهترین نمرات Chi-square انتخاب می‌شوند[1].
Random Forest (C)	هر درخت از جنگل تصادفی می‌تواند اهمیت یک ویژگی را باتوجه‌به توانایی خود در افزایش خلوص برگ‌ها محاسبه کند. این موضوع مربوط به نحوه عملکرد درختان طبقه‌بندی است. هر چه افزایش خلوص برگ‌ها بیشتر باشد، اهمیت ویژگی بالاتر است. این کار برای هر درخت انجام می‌شود، سپس در بین همه درختان میانگین گرفته می‌شود (Speiser et al., 2019).
Extra Tree (D)	این روش انتخاب ویژگی از بسیاری از جهات مشابه انتخاب ویژگی با استفاده از جنگل تصادفی است و تنها در نحوه ساخت درختان تصمیم‌گیری در جنگل تصادفی با آن متفاوت است. یکی از این تفاوت‌ها شیوه انتخاب نقاط انشعاب به‌منظور تقسیم گره‌ها است. جنگل تصادفی تقسیم بهینه را انتخاب می‌کند. درحالی‌که Extra tree این نقاط را به طور تصادفی انتخاب می‌کند (Geurts et al., 2006).

معیارهای ارزیابی مدل

از آنجا که طبقه‌بندی‌کننده‌های دودویی از محبوب‌ترین ابزارها هستند، فنون کارآمد بسیاری برای ارزیابی عملکرد آن‌ها وجود دارد. این معیارها به طور خلاصه در جدول ۵ آورده شده است. درجدول ۵، معیارهای ارزیابی مختلفی از جمله Accuracy، Recall، Precision و F1 ارائه شده است. برای تعیین اجزای معادلات این روش‌ها یعنی TP، TN، FN، و FP می‌توان از ماتریس درهم‌ریختگی مطابق جدول ۶ استفاده کرد. یکی از بهترین سنجه‌های ارزیابی عملکرد طبقه‌بندی‌کننده‌های دودویی، منحنی مشخصه عملکرد گیرنده (ROC[2]) است. این نمودار امکان مقایسه مثبت نادرست (پیش‌بینی نادرست بارش‌های سنگین، FP در جدول ۶) و مثبت درست (پیش‌بینی درست بارش‌های سنگین، TP در جدول ۶) را فراهم می‌کند. در نمودار ROC محور x مقادیر مثبت نادرست و محور y اندازه مقادیر مثبت درست هستند. یک روش قوی برای مقایسه مدل‌های طبقه‌بندی‌کننده مختلف، مقایسه منحنی‌های ROC آن‌ها است. برای مقایسه

1 https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223
2 Receiver operating characteristic curve

طبقه‌بندی‌کننده‌ها، ممکن است بخواهیم عملکرد ROC را به یک مقدار اسکالر واحد کاهش دهیم که عملکرد مورد انتظار را نشان می‌دهد. یک روش رایج، محاسبه مساحت زیر منحنی ROC است که به اختصار AUC[1] نامیده می شود (Hanley and McNeil, 1982; Bradley, 1997) . از آنجایی که AUC بخشی از مساحت مربع واحد است، مقدار آن همیشه بین ۰ تا ۱/۰ خواهد بود. با این حال، از آنجایی که حدس زدن تصادفی، خط مورب بین (۰، ۰) و (۱، ۱) را ایجاد می‌کند که مساحت آن ۰/۵ است، هیچ طبقه‌بندی واقعی نباید AUC کمتر از ۰/۵ داشته باشد (Fawcett, 2006).

جدول ۴- مدل‌های مورد استفاده و توضیح مربوط به آن‌ها

نام مدل	توضیحات
AdaBoost	طبقه‌بندی‌کننده AdaBoost یک برآوردکننده است که با برازش یک الگوریتم طبقه‌بندی‌کننده بر مجموعه داده اصلی شروع می‌شود و سپس کپی‌های اضافی از طبقه‌بندی‌کننده را در همان مجموعه داده قرار می‌دهد. در مجموعه جدید، وزن نمونه‌های طبقه‌بندی نادرست به‌گونه‌ای تنظیم می‌شوند که طبقه‌بندی‌کننده‌های بعدی بیشتر روی موارد دشوار تمرکز کنند (Freund and Schapire, 1997).
Decision Tree (DT)	الگوریتم‌های یادگیری مبتنی بر درخت یکی از متداول‌ترین روش‌های یادگیری نظارت شده است که با یادگیری قوانین اتخاذ شده از ویژگی‌ها، مقادیر پاسخ‌ها را پیش‌بینی می‌کند. گره ریشه (root node)، کل جمعیت را نشان می‌دهد، درحالی‌که گره‌های تصمیم‌گیری (decision nodes) نشان‌دهنده نقطه خاصی هستند که درخت تصمیم در مورد اینکه کدام ویژگی خاص تقسیم شود تصمیم می‌گیرد (Loh, 2011).
Naïve Bayes (NB)	NB یک تکنیک طبقه‌بندی بر اساس قضیه بیز است (Lindley, 1958). الگوریتم طبقه‌بندی‌کننده فرض می‌کند که یک ویژگی خاص در یک کلاس، به طور مستقیم با هیچ ویژگی دیگری در کلاس دیگر مرتبط نیست، اگرچه ویژگی‌های یک کلاس خاص می‌توانند بین خودشان وابستگی متقابل داشته باشند (Rish and Rish, 2001).
K Nearest Neighbor (KNN)	KNN یکی از ساده‌ترین و قدیمی‌ترین الگوریتم‌های طبقه‌بندی است. می‌توان KNN را نسخه ساده‌تری از طبقه‌بندی‌کننده NB تصور کرد. "K" در الگوریتم KNN تعداد نزدیکترین همسایه‌هایی است که برای گرفتن "رای" در نظر گرفته شده است. انتخاب مقادیر مختلف برای "K" می‌تواند نتایج طبقه‌بندی متفاوتی را برای یک شیء نمونه ایجاد کند (Cover and Hart, 1967).
Light GBM	Light GBM یک روش طبقه‌بندی با کارایی بالا بر اساس تقویت الگوریتم درخت تصمیم است. این روش برگ درخت را با بهترین تناسب تقسیم می‌کند، درحالی‌که سایر الگوریتم‌های تقویت‌کننده، درخت را از نظر عمق یا سطح تقسیم می‌کنند. از این رو منجر به دقت بسیار بهتری می‌شود که به ندرت می‌توان با هر یک از الگوریتم‌های تقویت‌کننده موجود به‌دست آورد. همچنین، به طور شگفت انگیزی بسیار سریع است و از این رو کلمه "Light" برای آن به کار برده می‌شود (Ke et al., 2017).
Logistic Regression (LR)	LR روشی قدرتمند و کاملاً شناخته شده برای طبقه‌بندی تحت نظارت است (Hosmer et al., 2013). رگرسیون لجستیک فرآیند مدل‌سازی احتمال یک نتیجه گسسته باتوجه‌به متغیر ورودی است. رایج ترین رگرسیون لجستیک یک نتیجه باینری را مدل می کند. چیزی که می‌تواند دو مقدار مانند true/false، yes/no و غیره داشته باشد (Edgar and Manz, 2017).
Neural Network (NN)	NN مجموعه‌ای از الگوریتم‌های یادگیری هستند که از عملکرد شبکه‌های عصبی مغز انسان الهام گرفته‌اند. آن‌ها ابتدا توسط McCulloch و Pitts پیشنهاد شدند (McCulloch and Pitts, 1943) و بعدها توسط آثار Rumelhart و همکاران در دهه ۱۹۸۰ رواج یافتند (Rumelhart et al., 1986). NN را می‌توان به‌عنوان یک گروه متصل از گره‌ها نشان داد. خروجی یک گره برای پردازش بعدی، باتوجه‌به اتصالات، به‌عنوان ورودی به گره دیگر می‌رود.
Random Forests (RF)	RF یک مدل قدرتمندتر است که ایده یک درخت تصمیم واحد را می‌گیرد و برای کاهش واریانس از صدها یا هزاران درخت، یک مدل جامع ایجاد می‌کند، همانند یک جنگل که مجموعه‌ای از بسیاری از درختان است. این الگوریتم از میانگین‌گیری برای بهبود دقت پیش‌بینی و کنترل برازش استفاده می‌کند، بنابراین مزیت به دست آوردن پیش‌بینی دقیق‌تر و پایدارتر را بدست می‌دهد (Breiman, 2001).
Support Vector Machine (SVM)	SVM می‌تواند داده‌های خطی و غیرخطی را طبقه‌بندی کند. ابتدا هر داده را در یک فضای n بعدی قرار می‌دهد که n تعداد ویژگی‌ها است. سپس خط جداکننده که داده‌ها را به دو کلاس جدا تقسیم می‌کند را در حالی شناسایی می‌کند که فاصله حاشیه‌ای را برای هر دو کلاس به حداکثر رسانده و خطاهای طبقه‌بندی را به حداقل برساند (Joachims, 1998).
XGBoost	تقویت درخت یک روش یادگیری ماشینی بسیار مؤثر و پرکاربرد است. XGBoost یک سیستم تقویت‌کننده درختی است که به طور گسترده توسط دانشمندان داده برای دستیابی به نتایج پیشرفته در بسیاری از چالش‌های یادگیری ماشین استفاده می‌شود (Chen and Guestrin, 2016).

در تحلیل پیش‌بینی در مسائل طبقه‌بندی، سنجه Accuracy برای مدل‌های پیش‌بینی، به‌تنهایی معیار مناسبی نیست. استفاده ساده از نتایج Accuracy می‌تواند گمراه‌کننده باشد و این سنجه اغلب معیار ضعیفی برای اندازه‌گیری عملکرد

1 Area Under Curve

است. بنابراین، توصیه شده است که هنگام ارزیابی مسائل تصمیم‌گیری دودویی، از منحنیROC استفاده شود (Provost and Fawcett, 1997, 1998; Provost et al., 1997). همچنین Precision و Recall در ارزیابی مسائل تصمیم‌گیری دودویی در مقایسه با Accuracy سنجه بهتری هستند (Valverde-Albacete et al., 2013). بنابراین، در این پژوهش نخستین سنجه ارزیابی برتری یک مدل نسبت به مدل دیگر سطح زیر منحنی (AUC) و پس از آن Precision است. Accuracy و F1-score نیز به‌عنوان سنجه‌های با اهمیت کمتر در جایگاه‌های بعدی قرار می‌گیرند.

جدول ۵- معیارهای ارزیابی و جزئیات آن‌ها (Novakovic et al., 2017)

معیار ارزیابی	فرمول	توضیحات
$Accuracy$	$\frac{TP+TN}{TP+TN+FP+FN}$	Accuracy به‌عنوان معیاری برای ارزیابی کیفیت مدل طبقه‌بندی است.
$Recall$	$\frac{TP}{TP+FN}$	Recall فقط به نحوه طبقه‌بندی نمونه‌های مثبت اهمیت می‌دهد. هنگامی که مدل، همه نمونه‌های مثبت را به‌عنوان مثبت طبقه‌بندی کند، Recall، صد درصد خواهد بود، حتی اگر همه نمونه‌های منفی به اشتباه به‌عنوان مثبت طبقه‌بندی شوند.
$Precision$	$\frac{TP}{TP+FP}$	Precision دقت مدل را در طبقه‌بندی یک نمونه به‌عنوان نمونه مثبت اندازه‌گیری می‌کند. وقتی مدل، طبقه‌بندی‌های مثبت نادرست زیاد، یا تعداد کمی طبقه‌بندی مثبت درست، ایجاد می‌کند؛ این عمل مخرج را افزایش داده و precision را کوچک می‌کند.
$F1$	$2\frac{P\times R}{P+R}$	F1 میانگین هارمونیک Precision و Recall است. هرچه ارزش نمره F1 به یک نزدیک‌تر باشد، عملکرد مورد انتظار مدل بهتر است.

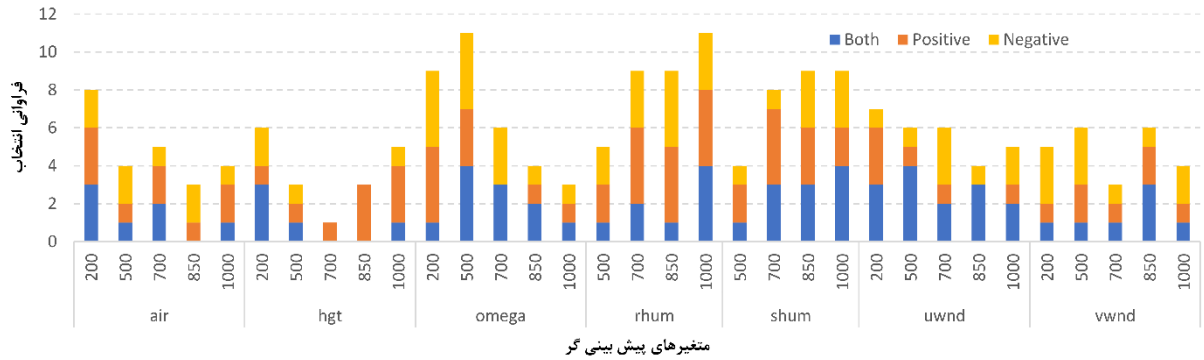
جدول ۶- ماتریس درهم‌ریختگی

نوع بارش پیش‌بینی شده	نوع بارش مشاهده شده	
	بارش سنگین	بارش غیر سنگین
بارش سنگین	بارش سنگین بوده و مدل آن را سنگین پیش‌بینی کرده است (TP)	بارش غیر سنگین بوده و مدل آن را سنگین پیش‌بینی کرده است (FP)
بارش غیر سنگین	بارش سنگین بوده و مدل آن را غیر سنگین پیش‌بینی کرده است (FN)	بارش غیر سنگین بوده و مدل آن را غیر سنگین پیش‌بینی کرده است (TN)

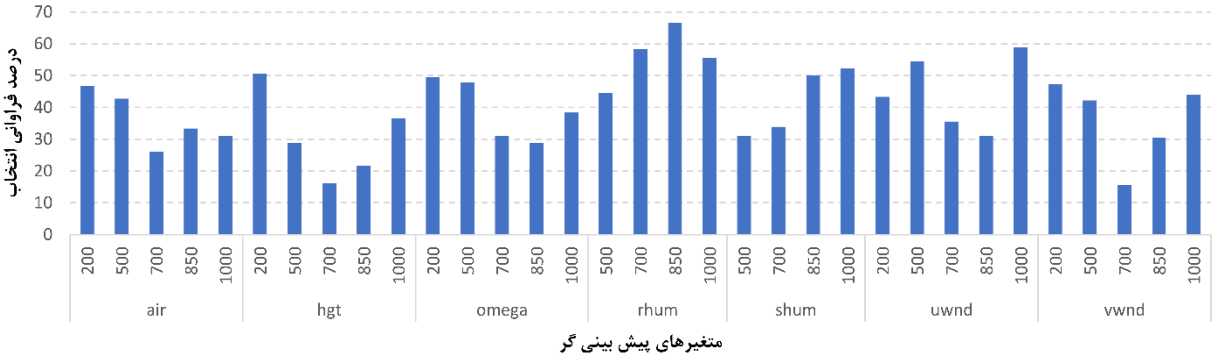
نتایج و بحث

انتخاب متغیرها

استقلال سی و سه متغیر پیش‌بینی گر تحت سه روش میانگین‌گیری (Negative، Positive، و Both) روی محدوده مطالعاتی با استفاده از چهار روش غربالگری مورد بررسی قرار گرفت. نتایج نشان داد که در همه روش‌های میانگین‌گیری، روش غربالگری A (یعنی Correlation) بیشترین تعداد متغیرها (۲۵-۲۸ متغیر) و روش غربالگری B (یعنی Chi-square) کمترین تعداد متغیرها (۱۰ متغیر) را در گام تأخیر زمانی یک روز قبل از بارش سنگین انتخاب کرده است (شکل ارائه نشده است). در مورد گام تأخیر زمانی یک روز قبل از بارش سنگین، فراوانی متغیرهای منتخب تحت سه روش میانگین‌گیری در شکل ۳ نشان داده شده است. متغیرهای rhum (تراز ۱۰۰۰ هکتوپاسکال) و omega (تراز ۵۰۰ هکتوپاسکال) بالاترین فراوانی انتخاب را در همه روش‌های غربالگری داشتند و نوع روش میانگین‌گیری از گریدها در انتخاب آن‌ها بی‌تاثیر بود. متغیرهای rhum (ترازهای ۷۰۰ و ۸۵۰ هکتوپاسکال)، و shum (ترازهای ۸۵۰ و ۱۰۰۰ هکتوسکال) و omega (تراز ۲۰۰ هکتوپاسکال) نیز فراوانی انتخاب قابل توجهی داشتند. کمترین فراوانی انتخاب را متغیر hgt (تراز ۷۰۰ هکتوپاسکال) داشت که تنها در روش میانگین‌گیری از گریدهای مثبت این انتخاب انجام شد.



شکل ۳- فراوانی انتخاب متغیرهای پیش‌بینی گر تحت سه روش میانگین‌گیری از شبکه گرید برای مدل‌سازی بر مبنای داده‌های یک روز پیش از رخداد بارش سنگین



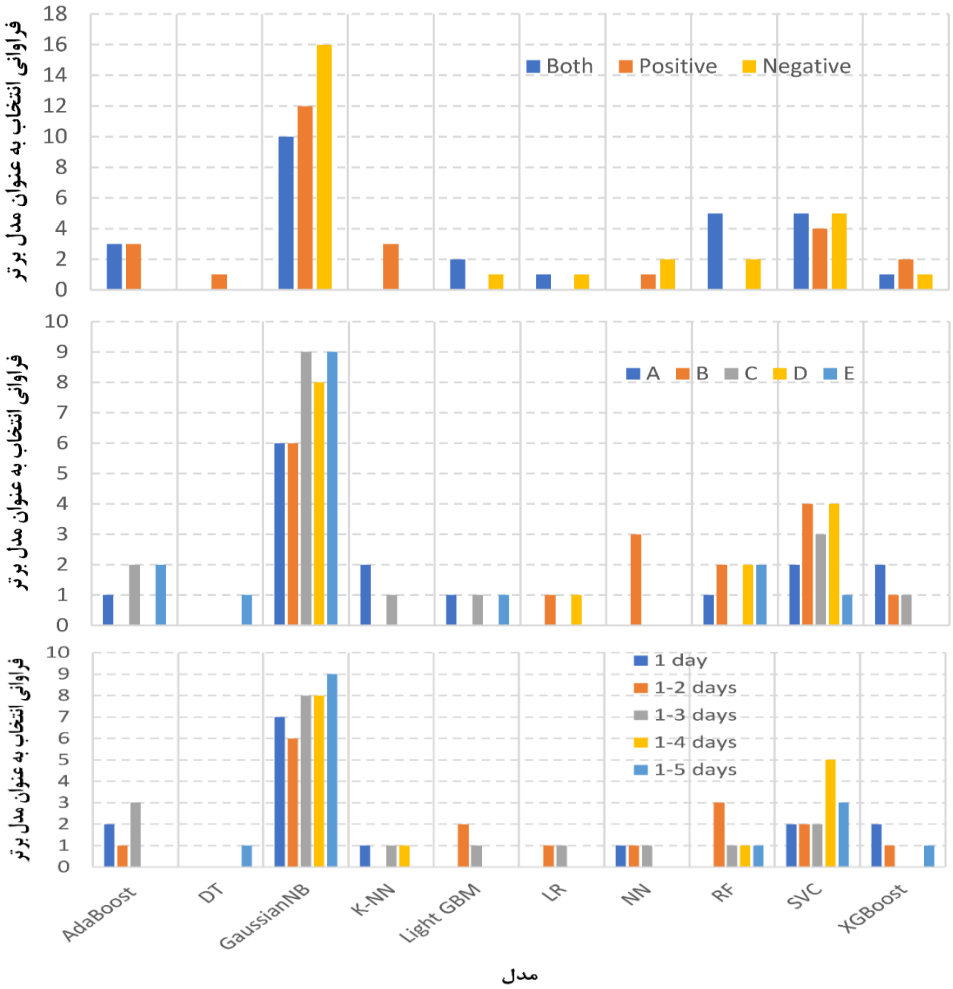
شکل ۴- درصد فراوانی انتخاب متغیرهای پیش‌بینی گر

محاسبات مربوط به انتخاب متغیرها، برای ۱-۲ روز، ۱-۳ روز، ۱-۴ روز، و ۱-۵ روز قبل از بارش سنگین به طور مستقل انجام شد. درصد فراوانی انتخاب هر یک از متغیرهای پیش‌بینی گر در سرجمع روش میانگین‌گیری از گریدها، تأخیر زمانی و روش انتخاب متغیر در شکل ۴ ارائه شده است. باتوجه‌به شکل مذکور، تراز ۲۰۰ هکتوپاسکال متغیرهای دما (air)، ارتفاع ژئوپتانسیل (hgt)، امگا (omega) و مولفه نصف‌النهاری باد (vwnd)، تراز ۵۰۰ هکتوپاسکال مولفه مداری باد (uwnd)، تراز ۸۵۰ هکتوپاسکال رطوبت نسبی (rhum) و تراز ۱۰۰۰ هکتوپاسکال مولفه افقی سرعت برداری باد (uwnd) و رطوبت ویژه (shum) از بیشترین فراوانی انتخاب در هر گروه از متغیرهای هم تراز برخوردارند. بیشترین درصد فراوانی انتخاب (۶۷ درصد) مربوط به رطوبت نسبی تراز ۸۵۰ هکتوپاسکال و کمترین درصد فراوانی انتخاب مربوط به ارتفاع ژئوپتانسیل و مولفه نصف‌النهاری باد تراز ۷۰۰ هکتوپاسکال (۱۶ درصد) است.

تأثیر روش میانگین‌گیری، روش انتخاب متغیر و گام تأخیر در انتخاب مدل‌های برتر پیش‌بینی

شکل ۵ فراوانی انتخاب هر یک از مدل‌های داده‌کاوی به‌عنوان مدل برتر تحت تأثیر روش میانگین‌گیری، روش انتخاب متغیرها و گام تأخیر زمانی متغیرها را نشان می‌دهد. باتوجه‌به شکل مذکور نتیجه گرفته می‌شود: الف) در بین سه روش میانگین‌گیری از گریدهای بی‌هنجاری شدید، روش Negative بالاترین فراوانی مدل‌های برتر را دارد. در هر سه روش میانگین‌گیری، مدل‌های GaussianNB و SVC بیشتر از دیگر مدل‌ها به‌عنوان مدل برتر انتخاب شده‌اند (شکل ۵-بالا). ب) با استفاده از همه روش‌های انتخاب متغیر، مدل GaussianNB بیشتر از سایر مدل‌ها به‌عنوان مدل برتر معرفی شده است و مدل SVC در جایگاه دوم قرار می‌گیرد. فراوانی انتخاب مدل GaussianNB به‌عنوان مدل برتر پیش‌بینی با استفاده از روش انتخاب متغیر C بیشتر از دیگر روش‌های انتخاب متغیر است. با این وجود، بین روش برتر انتخاب متغیر C و استفاده از همه متغیرها (روش E) از نظر فراوانی انتخاب مدل GaussianNB به‌عنوان مدل برتر تفاوتی وجود ندارد. البته در مورد مدل SVC تعداد دفعاتی که بدون غربال متغیرها به‌عنوان مدل برتر انتخاب شده کمتر از روش‌های غربال متغیرها است. باید اضافه شود که غربال متغیرها تاثیری در انتخاب مدل DT به‌عنوان مدل برتر نداشته است و حتی در یک مورد عدم غربال متغیرها باعث برتری این مدل نسبت به دیگر مدل‌ها شده است. (شکل ۵-وسط). ج) در صورتی که گام تأخیر زمانی متغیرها در پیش‌بینی بارش سنگین اهمیت داشته باشد، فراوانی انتخاب مدل‌های GaussianNB و SVC بالاتر از دیگر مدل‌ها است. بر مبنای GaussianNB گام تأخیر زمانی تا ۵ روز و براساس مدل

SVC گام تأخیر زمانی تا ۴ روز، فراوانی بالاتری نسبت به دیگر گام‌های تأخیر دارند. در اینجا نیز مدل DT ضعیف‌ترین نتایج را تولید نموده و تنها یک‌بار در گام تأخیر زمانی تا پنج روز به‌عنوان مدل برتر شناخته شده است (شکل ۵-پایین). نتیجه مهمی که از مقایسه سه نمودار شکل ۵ به‌دست می‌آید برتری آشکار مدل GaussianNB بر دیگر مدل‌ها از هر سه جنبه یاد شده است و مدل SVC در جایگاه دوم قرار می‌گیرد.



شکل ۵- فراوانی مدل‌های برتر برای پیش‌بینی بارش سنگین بر مبنای روش میانگین‌گیری متغیرها (بالا)، روش انتخاب متغیرها (وسط) و گام تأخیر زمانی متغیرها (پایین)

جدول ۷- برترین مدل‌های پیش‌بینی بارش سنگین با توجه به گام‌های تاخیر زمانی متغیرهای پیش‌بینی گر

	گام تأخیر زمانی				
	1 day	1 - 2 days	1 - 3 days	1 - 4 days	1 - 5 days
برترین مدل	GaussianNB	LR	LR	RF	SVC
روش میانگین‌گیری	Both	Negative	Both	Both	Negative
روش انتخاب متغیر	B	D	B	B	D
AC	۰/۷۸۳	۰/۸۰۴	۰/۷۶۱	۰/۸۴۸	۰/۸۲۶
AUC	۰/۷۸۱	۰/۸۰۵	۰/۷۵۷	۰/۸۴۹	۰/۸۲۱
TP	۲۰	۲۰	۲۰	۲۱	۲۲
FP	۵	۵	۵	۴	۳
FN	۶	۴	۶	۳	۵
TN	۱۵	۱۷	۱۵	۱۸	۱۶
P	۰/۸۰۰	۰/۸۳۳	۰/۷۶۹	۰/۸۷۵	۰/۸۱۵
F1	۰/۸۰۰	۰/۸۱۶	۰/۷۸۴	۰/۸۵۷	۰/۸۴۶

معرفی برترین مدل پیش‌بینی

در جدول ۷ برترین مدل‌های پیش‌بینی در نتیجه اجرای ده مدل داده‌کاوی با چهار روش انتخاب متغیر تحت سه روش میانگین‌گیری متغیرها ارائه شده است. در بخش قبل، فراوانی انتخاب مدل‌های برتر ارائه شد. در اینجا، هدف معرفی مدلی است که به لحاظ سنجه‌های مختلف ارزیابی، بالاترین کارایی را به‌دست آورده است. باتوجه به جدول ۷ مشاهده می‌شود که هنگامی که از داده‌های یک روز قبل در پیش‌بینی استفاده شود، مدل GaussianNB بالاترین کارایی را از نظر سنجه‌های AUC، AC، P و F1 داشته است. هنگامی که ورودی به مدل‌ها داده‌های ۱-۲ و ۱-۳ روز پیش از رخداد بارش سنگین باشد، مدل LR برترین مدل پیش‌بینی خواهد بود. در حالت استفاده از داده‌های ۱-۴ و ۱-۵ روز پیش از وقوع بارش سنگین، به ترتیب، مدل‌های RF و SVC مدل‌های برتر خواهند بود. نکته قابل توجه از جدول ۷ این است که روش میانگین‌گیری Both (هر دوی گریدهای بی‌هنجاری شدید مثبت و منفی) و Negative (گریدهای بی‌هنجاری شدید منفی) همراه با روش‌های غربال متغیر B و D در انتخاب مدل‌های برتر نقش داشته‌اند.

با مقایسه مدل‌های برتر در جدول ۷ می‌توان نتیجه گرفت که مدل RF (جنگل تصادفی) با داده‌های ورودی ۱-۴ روز قبل، روش میانگین‌گیری Both و روش غربالگری B از نظر سنجه‌های AC، AUC، P و F1 (به ترتیب با مقادیر ۰/۸۴۸، ۰/۸۴۹، ۰/۸۷۵ و ۰/۸۵۷) بهترین مدل طبقه‌بندی در تشخیص بارش‌های سنگین از غیر سنگین است. طبق جدول ۷، این مدل توانسته است در مرحله صحت‌سنجی، ۲۱ مورد از ۲۵ رخداد بارش سنگین را به‌درستی پیش‌بینی کند. متغیرهای مورداستفاده در برترین مدل عبارت‌اند از: uwnd (ترازهای ۱۰۰۰ و ۸۵۰ هکتوپاسکال) در چهار روز قبل، uwnd (تراز ۷۰۰ هکتوپاسکال) در سه روز قبل؛ uwnd (ترازهای ۵۰۰، ۷۰۰، ۸۵۰ و ۱۰۰۰ هکتوپاسکال) و rhum (تراز ۱۰۰۰ هکتوپاسکال) در دو روز قبل؛ shum (ترازهای ۱۰۰۰ و ۸۵۰ پاسکال) در یک روز قبل از رخداد بارش سنگین.

نتیجه‌گیری

در این پژوهش از چهار روش انتخاب متغیر و ده مدل یادگیری ماشین از نوع طبقه‌بندی‌کننده دودوئی، به‌منظور مدل‌سازی و پیش‌بینی بارش‌های سنگین منطقه‌ای استفاده شد. داده‌های متغیرهای همدیدی تا پنج روز پیش از رخداد بارش سنگین به‌عنوان ورودی این مدل‌ها مورداستفاده قرار گرفت. پس از آن تحلیل مقایسه‌ای به‌منظور تعیین بهترین مدل برای پیش‌بینی بارش‌های سنگین انجام شد. بر اساس تحلیل فنون داده کاوی، مدل طبقه‌بندی جنگل تصادفی (RF) با داده‌های ورودی ۱-۴ روز قبل، بالاترین کارایی را از نظر سنجه‌های AUC، AC، P و F1 در تشخیص بارش‌های سنگین از غیر سنگین داشت. متغیرهای rhum، shum و uwnd مهمترین عوامل در پیش‌بینی بارش‌های سنگین بود. در پیش‌بینی بارش ایستگاه همدید کرمانشاه (Omidvar et al., 2014) و در پیش‌بینی بارش ایستگاه ساری (Baharian and Salimi, 2018)، مدل درخت تصمیم را به‌عنوان مدلی کارآمد در این زمینه معرفی کرده‌اند. آن‌چه از نتایج این پژوهش حاصل شد، نشان می‌دهد که بر خلاف یافته‌های دو پژوهش اشاره شده، که از معدود مطالعات انجام شده در رابطه با پیش‌بینی بارش ایران با استفاده از روش‌های داده‌کاوی می‌باشند، مدل درخت تصمیم مدلی بسیار ضعیف در تشخیص بارش‌های سنگین از غیر سنگین بوده و در هیچ یک از مراحل اجرای سناریو به‌عنوان مدل برتر از بین ده مدل یادگیری ماشین انتخاب نشده است.

هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد.

REFERENCES

Abbot J., Marohasy J. (2014) Input selection and optimisation for monthly rainfall forecasting in queensland, australia, using artificial neural networks. Atmos Res 138:166–178. https://doi.org/10.1016/j.atmosres.2013.11.002

Aftab S., Ahmad M., Hameed N., Bashir M.S., Ali I., Nawaz Z. (2018) Rainfall prediction in Lahore City using data mining techniques. Int J Adv Comput Sci Appl 9:254–260. https://doi.org/10.14569/IJACSA.2018.090439

Ahmad M., Aftab S. (2017) Analyzing the performance of svm for polarity detection with different datasets. Int J Mod Educ Comput Sci 9:29–36. https://doi.org/10.5815/ijmecs.2017.10.04

Ahmad M., Aftab S., Ali I. (2017) Sentiment analysis of tweets using SVM. Int J Comput Appl 177:25–29. https://doi.org/10.5120/IJCA2017915758

Ahmad M., Aftab S., Ali I., Hameed N. (2017) Hybrid tools and techniques for sentiment analysis: a review. Int J Multidiscip Sci Eng 8:28–33

Ahmad M., Aftab S., Muhammad S.S., Ahmad S. (2017) Machine learning techniques for sentiment analysis: a review. Int J Multidiscip Sci Eng 8:27–32

Alijani B., O'Brien J., Yarnal B., O'Brien J., Yarnal B. (2008) Spatial analysis of precipitation intensity

and concentration in Iran. Theor Appl Climatol 94:107–124. https://doi.org/10.1007/s00704-007-0344-y

Arvin A., Mohamadinejad J. (2015) Synoptic survey of floods caused by heavy rainfall of 4 february 2006 in the lorestan basin. J Nat Environ Hazards 4:75-90 (In Farsi)

Baharian A., Salimi A. (2018) Utilizing of decision tree model in predicting precipitation in Sari based on the information from Sari synoptic station. In: The first national conference on management strategies of water resources and environmental challenges. pp 1-10 (In Farsi)

Beguería S., Angulo-Martínez M., Vicente-Serrano S.M., López-Moreno J.I., El-Kenawy A. (2011) Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. Int J Climatol 31:2102–2114. https://doi.org/10.1002/JOC.2218

Bradley A.P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 30:1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman L. (2001) Random Forests. Mach Learn 2001 451 45:5–32. https://doi.org/10.1023/A:1010933404324

Cavazos T., Turrent C., Lettenmaier D.P. (2008) Extreme precipitation trends associated with tropical cyclones in the core of the North American monsoon. Geophys Res Lett 35:L21703. https://doi.org/10.1029/2008GL035832

Chen T., Guestrin C. (2016) XGBoost: A Scalable Tree Boosting System. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 13-17-Augu:785–794. https://doi.org/10.1145/2939672.2939785

Edgar T.W., Manz D.O. (2017) Exploratory Study. Res Methods Cyber Secur 95–130. https://doi.org/10.1016/B978-0-12-805349-2.00004-2

Fawcett T. (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874. https://doi.org/10.1016/J.PATREC.2005.10.010

Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (1996) Advances in knowledge discovery and data mining. American Association for Artificial Intelligence Menlo Park, CA, USA ©1996

Freund Y., Schapire R.E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139. https://doi.org/10.1006/JCSS.1997.1504

Geurts P., Ernst D., Wehenkel L. (2006) Extremely randomized trees. Mach Learn 2006 631 63:3–42. https://doi.org/10.1007/S10994-006-6226-1

Groisman P.Y., Knight R.W., Easterling D.R., Karl T.R., Hegerl G.C., Razuvaev V.N. (2005) Trends in intense precipitation in the climate record. J Clim 18:1326–1350. https://doi.org/10.1175/JCLI3339.1

Gupta A., Farhan Habib M., Mandal U., Chowdhury P., Tornatore M., Mukherjee B. (2018) On service-chaining strategies using virtual network functions in operator networks. Comput Networks 133:1–16. https://doi.org/10.1016/j.comnet.2018.01.028

Hall M.A. (2000) Correlation-based feature selection of discrete and numeric class machine learning

Hanley J.A., McNeil B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36. https://doi.org/10.1148/RADIOLOGY.143.1.7063747

Hirsch R.M., Archfield S.A. (2015) Flood trends: Not higher but more often. Nat Clim Chang 5:198–199. https://doi.org/10.1038/NCLIMATE2551

Hosmer D.W., Lemeshow S., Sturdivant R.X. (2013) Applied logistic regression: third edition. Appl Logist Regres Third Ed 1–510. https://doi.org/10.1002/9781118548387

IPCC (2007) IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.[Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and

IPCC (2012) IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance ClimateChange Adaptation. A Special Report of Working Groups I and II of theIntergovernmental Panel on Climate Change [; Field, C.B., V. Barros, T.F. Stocker,D. Qin, D.J. Dokken, K.

Jafar Nazemosadat M., Shahgholian K. (2017) Heavy precipitation in the southwest of Iran: association with the Madden–Julian Oscillation and synoptic scale analysis. Clim Dyn 49:3091–3109. https://doi.org/10.1007/s00382-016-3496-6

Joachims T. (1998) Making large-scale SVM learning practical

Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.Y. (2017) LightGBM: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017-Decem:3147–3155

Khalili A., Rahimi J. (2018) Climate. In: Roozitalab MH, Siadat H, Farshad A (eds) The Soils of Iran. Springer International Publishing, Cham, pp 19–33

Khoshakhlagh F., Safaierad R., Salmani D. (2015) The Synoptic analysis of flood occurrence on November 2011 in Behbahan and Likak cities. Phys Geogr Res 46:509-523 (In Farsi). https://doi.org/10.22059/JPHGR.2014.53001

Lindley D. V. (1958) Fiducial distributions and bayes' theorem. J R Stat Soc Ser B 20:102–107. https://doi.org/10.1111/J.2517-6161.1958.TB00278.X

Loh W.Y. (2011) Classification and regression trees. Wiley Interdiscip Rev Data Min Knowl Discov 1:14–23. https://doi.org/10.1002/WIDM.8

Longadge R., Dongre S.S., Malik L. (2013) Class Imbalance Problem in Data Mining: Review. Int J

Comput Sci Netw 2:

Mallakpour I., Villarini G. (2015) The changing nature of flooding across the central United States. Nat Clim Chang 2014 53 5:250–254. https://doi.org/10.1038/nclimate2516

McCulloch W.S., Pitts W. (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943 54 5:115–133. https://doi.org/10.1007/BF02478259

Mishra N., Soni H.K., Sharma S., Upadhyay A.K. (2017) A comprehensive survey of data mining techniques on time series data for rainfall prediction. J ICT Res Appl 11:167–183. https://doi.org/10.5614/ITBJ.ICT.RES.APPL.2017.11.2.4

Nayak D.R., Mahapatra A., Mishra P., RanjanNayak D., Mahapatra A., Mishra P. (2013) A survey on rainfall prediction using artificial neural network. Int J Comput Appl 72:32–40. https://doi.org/10.5120/12580-9217

Nayak M.A., Ghosh S. (2013) Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. Theor Appl Climatol 114:583–603. https://doi.org/10.1007/S00704-013-0867-3/TABLES/9

Novakovic J., Veljovi A., Iiic S., Papic Z., Tomovic M. (2017) Evaluation of classification models in machine learning. Theory Appl Math Comput Sci 7:39–46

Omidvar K., Shafie S., Taghizadeh Z., Alipoor M. (2014) Assessing the performance of decision tree model in predicting precipitation in kermanshah synoptic station. J Appl Res Geogr Sci 14:89-110 (In Farsi)

Pourasghar F., Oliver E.C.J., Holbrook N.J. (2021) Influence of the MJO on daily surface air temperature over Iran. Int J Climatol 41:4562–4573. https://doi.org/10.1002/JOC.7086

Poursalehi F., Shahid A., Khasheisiuk A. (2019) Comparison of decision tree m5 and k-nearest neighborhood algorithm models in the prediction of monthly precipitation (case study: birjand synoptic station). Iran J Irrig Drain 13:1283-1293 (In Farsi)

Provost F., Fawcett T. (1997) Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. Proc THIRD Int Conf Knowl Discov DATA Min 43--48

Provost F., Fawcett T. (1998) Robust classiication systems for imprecise environments. Proc AAAI-98 AAAI Press Menlo Park CA 706–713

Provost F., Fawcett T., Kohavi R. (1997) The case against accuracy estimation for comparing induction algorithms. Proc FIFTEENTH Int Conf Mach Learn 445--453

Rahimi M., Fatemi S.S. (2019) Mean versus extreme precipitation trends in iran over the period 1960–2017. Pure Appl Geophys 2019 1768 176:3717–3735. https://doi.org/10.1007/S00024-019-02165-9

Rish I., Rish I. (2001) An empirical study of the naive bayes classifier

Ruivo H.M., De Campos Velho H.F., Sampaio G., Ramos F.M. (2015) Analysis of extreme precipitation events using a novel data mining approach. Am J Environ Eng 5:96–105. https://doi.org/10.5923/s.ajee.201501.13

Rumelhart D.E., Hinton G.E., Williams R.J. (1986) Learning representations by back-propagating errors. Nat 1986 3236088 323:533–536. https://doi.org/10.1038/323533a0

Seneviratne S.I., Nicholls N., Easterling D., Goodess C.M., Kanae S., Kossin J., Luo Y., Marengo J., Mc Innes K., Rahimi M., Reichstein M., Sorteberg A., Vera C., Zhang X., Rusticucci M., Semenov V., Alexander L. V., Allen S., Benito G., Cavazos T., Clague J., Conway D., Della-Marta P.M., Gerber M., Gong S., Goswami B.N., Hemer M., Huggel C., Van den Hurk B., Kharin V. V., Kitoh A., Klein Tank A.M.G., Li G., Mason S., Mc Guire W., Van Oldenborgh G.J., Orlowsky B., Smith S., Thiaw W., Velegrakis A., Yiou P., Zhang T., Zhou T., Zwiers F.W. (2012) Changes in climate extremes and their impacts on the natural physical environment. In Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Cl

Speiser J.L., Miller M.E., Tooze J., Ip E. (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 134:93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Sun C., Huang G., Fan Y. (2020) Multi-indicator evaluation for extreme precipitation events in the past 60 years over the Loess Plateau. Water (Switzerland) 12:. https://doi.org/10.3390/w12010193

Vaghefi S.A., Keykhai M., Jahanbakhshi F., Sheikholeslami J., Ahmadi A., Yang H., Abbaspour K.C. (2019) The future of extreme climate in Iran. Sci Reports 2019 91 9:1–11. https://doi.org/10.1038/s41598-018-38071-8

Valverde-Albacete F.J., Carrillo-de-Albornoz J., Peláez-Moreno C. (2013) A proposal for new evaluation metrics and result visualization technique for sentiment analysis tasks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg, pp 41–52

Wheater H.S. (2002) Progress in and prospects for fluvial flood modelling. Philos Trans R Soc London Ser A Math Phys Eng Sci 360:1409–1431. https://doi.org/10.1098/rsta.2002.1007

WMO (2016) Guidelines on the defintion and monitoring of extreme weather and climate events. Task Team Defin Extrem Weather Clim Events, WMO, 4/14/2016 62. https://doi.org/10.1109/CSCI.2015.171

Young P.C. (2002) Advances in real–time flood

https://doi.org/10.18517/IJASEIT.6.6.1487
Zhang S., Lu L., Yu J., Zhou H. (2016) Short-term water level prediction using different artificial intelligent models. 2016 5th Int Conf Agro-Geoinformatics, Agro-Geoinformatics 2016. https://doi.org/10.1109/AGRO-GEOINFORMATICS.2016.7577678
forecasting. Philos Trans R Soc London Ser A Math Phys Eng Sci 360:1433–1450. https://doi.org/10.1098/rsta.2002.1008
Zainudin S., Jasim D.S., Bakar A.A. (2016) Comparative analysis of data mining techniques for malaysian rainfall prediction. Int J Adv Sci Eng Inf Technol 6:1148–1153.