



## Evaluating and Forecasting the Probability of Lightning Occurrence in Rasht City

Afsaneh Ghasemi<sup>1\*</sup>, Jamil Amanollahi<sup>1</sup>

<sup>1</sup> Department of Environmental Sciences and Engineering, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

### ARTICLE INFO

Article Type: Research article

Article history:

Received 20 October 2019

Accepted 16 February 2020

Available online 11 March 2020

Keywords:

Lightning, Prediction, Rasht, Data Mining Models, Comparative Criteria.

Citation: Ghasemi, A., Amanollahi, J. (2020). Evaluating and Forecasting the Probability of Lightning Occurrence in Rasht City. *Geography and Sustainability of Environment*, 10 (1), 21-35.

doi: [10.22126/GES.2020.4534.2122](https://doi.org/10.22126/GES.2020.4534.2122)

### ABSTRACT

Lightning is one of the most severe weather hazards that will cause significant economic, social and environmental damage each year. The prediction of a lightning is a very difficult task due to the spatial and temporal expansion of weather either physically or dynamically. Therefore, timely forecasting of lightning and evaluation of the best data mining model is effective in reducing damage. In this research, the data of the years 2012\_2018 of the Meteorological Station of Rasht were used, including dependent variable of occurrence and non-occurrence of lightning during 7 years and independent variables of factors affecting lightning including temperature, relative humidity, cloudy, wind speed, wind direction, pressure air and Previous day's lightning. After preprocessing and processing data, data mining models including Classification & Regression Tree (CART), Chi-squared Automatic Interaction Detector (CHAID), Induction of Decision Trees (C5) and neural networks Radial Basis Function (RBF), Multi Layer Perceptron (MLP) and Support Vector Machine (SVM) were used in SPSS Modeler Ver 20 software. The results of the models were compared with the Comparative Criteria and the Receiver operating characteristic (ROC) curve. According to the results of the models, the probability of lightning occurrence is higher in the months of May, June and July than in other months and the rate of occurrence from spring to winter has a decreasing trend, while in winter it is at least. CHAID tree with a specificity rate of 0.794 and a minimum false positive rate of 0.205 and the SVM model with a correct prediction of 0.773 and an error rate of 0.475 and precision of 0.855 have optimum performance compared with other models.



## ارزیابی و پیش‌بینی احتمال وقوع رعد و برق در شهر رشت

افسانه قاسمی<sup>۱\*</sup>، جمیل امان‌اللهی<sup>۱</sup>

<sup>۱</sup> گروه علوم و مهندسی محیط‌زیست، دانشکده منابع طبیعی، دانشگاه کردستان، سنندج، ایران

### چکیده

رعد و برق از شدیدترین مخاطرات آب‌وهوایی است که هر ساله به خسارت اقتصادی - اجتماعی و زیست‌محیطی فراوانی منجر می‌شود. پیش‌بینی رعد و برق به‌علاوه گسترش فضایی و زمانی آب‌وهوا به‌صورت فیزیکی یا دینامیکی بسیار دشوار است؛ بنابراین پیش‌بینی به‌موقع و ارزیابی بهترین مدل داده‌کاوی در کاهش آسیب و خسارت مؤثر است. در پژوهش حاضر، از داده‌های سال ۱۳۹۰ تا ۱۳۹۶ ایستگاه هواشناسی رشت استفاده شد. متغیر وابسته وقوع و عدم وقوع رعد و برق در طی هفت سال و متغیرهای مستقل عوامل مؤثر بر رعد و برق شامل دما، رطوبت نسبی، ابرناکی، سرعت باد، جهت باد، فشار هوا و رعد و برق در روز گذشته است. پس از پیش‌پردازش و پردازش داده‌ها از مدل‌های داده‌کاوی شامل درخت‌های کارت، چاید، سی فایو و شبکه عصبی پرسپترون چندلایه، تابع پایه‌ای شعاعی و ماشین بردار پشتیبان در نرم‌افزار اس.اس.پی.اس. مودلر ورژن ۲۰ استفاده شد. نتایج حاصل از مدل‌ها با معیارهای مقایسه‌ای و منحنی راک مقایسه شد. با توجه به نتایج به‌دست‌آمده از مدل‌ها، احتمال وقوع رعد و برق در آینده در ماه‌های اردیبهشت، خرداد و تیر نسبت به سایر ماه‌ها حداکثر است و میزان وقوع از فصل بهار تا زمستان روند کاهشی دارد و در فصل زمستان، به حداقل مقدار خود می‌رسد و از میان مدل‌های پیش‌بینی‌کننده، درخت چاید با میزان تشخیص ۰/۷۹۴ و حداقل میزان نرخ مثبت کاذب ۰/۲۰۵ و مدل ماشین بردار پشتیبان با پیش‌بینی صحیح ۰/۷۷۳ مورد و نرخ خطای ۰/۴۷۵ و دقت ۰/۸۵۵ نسبت به سایر مدل‌ها عملکرد بهینه دارند.

### مشخصات مقاله

نوع مقاله: پژوهشی

تاریخچه مقاله:

دریافت ۲۸ مهر ۱۳۹۸

پذیرش ۲۷ بهمن ۱۳۹۸

دسترسی آنلاین ۲۱ اسفند ۱۳۹۸

کلیدواژه‌ها:

رعد و برق، پیش‌بینی، رشت، مدل‌های داده‌کاوی، معیارهای مقایسه‌ای.

/استناد: قاسمی، افسانه؛ امان‌اللهی، جمیل (۱۳۹۹). ارزیابی و پیش‌بینی احتمال وقوع رعد و برق در شهر رشت. *جغرافیا و پایداری محیط*، ۱۰ (۱)، ۲۱-۳۵.

doi: [10.22126/GES.2020.4534.2122](https://doi.org/10.22126/GES.2020.4534.2122)

## مقدمه

رعد و برق، تخلیه الکتریکی سریع و ناشی از شرایط جوئی است. این تخلیه الکتریسیته موجب فوران ناگهانی موج‌های نور و صدا می‌شود؛ در واقع افزایش اختلاف پتانسیل بین ابر و زمین سبب یونیزه شدن جو و ایجاد موج سریع و شدیدی در هوای اطراف می‌شود که خود را به صورت رعد و برق نشان می‌دهد. رطوبت و جو ناپایدار نقش مهمی در تشکیل رعد و برق دارند، تجربیات آزمایشگاهی و آزمایش‌های عددی نشان می‌دهد که لایه فاز اختلاط ابر که شامل برف، یخ و آب با دمای بسیار پایین است، شرایط ماکروفیزیکی لازم برای وقوع رعد و برق را فراهم می‌سازد (کوداما<sup>۱</sup> و همکاران، ۲۰۰۷). از آنجاکه فعالیت‌های رعد و برقی در نزدیکی منبع آبی و کوه‌های مرطوب افزایش می‌یابد (هودانیش و وولین<sup>۲</sup>، ۲۰۱۲) مشخص کردن آن در مناطق مختلف جغرافیایی برای پژوهش‌ها و برنامه‌های پیش‌بینی شده بسیار مهم است (بارنس و نیوتن<sup>۳</sup>، ۱۹۸۲: ۱۴۷؛ کورت و گریفیث<sup>۴</sup>، ۱۹۸۲: ۳۴)؛ زیرا افزایش شدت رعد و برق می‌تواند سبب ایجاد توفان‌های تدری، بادهای قوی، باران شدید، سیلاب و گردبادها شود (راجیوان<sup>۵</sup> و همکاران، ۲۰۱۲).

در طول سال‌های اخیر برخی از مدل‌های آماری مورد استفاده به منظور کاهش خسارت ناشی از رعد و برق شامل رگرسیون خطی چندگانه، رگرسیون لجستیک باینری، طبقه‌بندی و درخت رگرسیون (کارت<sup>۶</sup>) است (لامبرت<sup>۷</sup> و همکاران، ۲۰۰۵)؛ در واقع این روش‌ها ارتباط بین مجموعه‌ای از پیش‌بینی‌ها و احتمال توفان‌های تدری یا فرکانس رعد و برق را تعیین می‌کنند (ریپ<sup>۸</sup>، ۱۹۹۴). تاکنون پژوهش‌های جهانی بسیاری برای پیش‌بینی رعد و برق و تأثیر متغیرهای گوناگون انجام شده که همگی با روش‌های متعدد و مدل‌سازی به منظور افزایش صحت و دقت در پیش‌بینی و کاهش آسیب و خسارت ناشی از رعد و برق صورت گرفته است؛ برای نمونه به منظور پیش‌بینی رعد و برق در جنوب شرقی هند از پارامترهای رطوبت، دما، بی‌ثباتی جو و ۱۱۵ مورد رعد و برق در طول دوره زمانی ۲۰۰۶ تا ۲۰۰۹ برای توسعه مدل آماری مبتنی بر رگرسیون استفاده شد.

نتایج نهایی نشان داد مدل پیش‌بینی دقت قابل قبولی دارد؛ اما براساس نتایج به‌دست‌آمده از تجزیه و تحلیل‌های آماری بهبود در پیش‌بینی وقوع رعد و برق و توفان‌های تدری و کاهش نرخ اشتباه در مدل پیش‌بینی به میزان و تأثیر متغیرهای وارد شده به مدل و مشاهدات بیشتر در مناطقی که فرکانس رعد و برق بیشتری دارند، وابسته است و در صورت وجود متغیرهای بیشتر می‌توان پیش‌بینی را برای فصل‌های مختلف به‌طور جداگانه و با دقت بیشتر انجام داد (راجیوان و همکاران، ۲۰۱۲)؛ در این راستا به منظور تحلیل روند وقوع رعد و برق از داده‌های هواشناسی هفده ایستگاه سینوپتیک با طولانی‌ترین دوره آماری و دو آزمون ناپارامتری استفاده شد و براساس نتایج به‌دست‌آمده، مشخص شد نیمه غربی ایران از نظر تعداد وقوع روزهای همراه با رعد و برق، منطقه همگنی نیست و تعداد وقوع به‌صورت سالیانه از شمال به جنوب کاهش می‌یابد و در مقیاس فصلی و سالیانه روند افزایشی دارد (رسولی و جوان، ۱۳۹۱)؛ همچنین با کدهای مربوط به توفان‌های تدری، فراوانی و توزیع زمانی و مکانی توفان‌های تدری در شمال غربی ایران طی سال‌های ۱۹۵۱ تا ۲۰۰۲ تحلیل و مشخص شد که فصل‌های بهار و تابستان، حداکثر بارش رعد و برقی را دارند (رسولی و همکاران، ۱۳۸۶).

1- Kodama

2- Hodanish &amp; Wolyn

3- Barnes &amp; Newton

4- Court &amp; Griffiths

5- Rajeevan

6- Classification &amp; Regression Tree (CART)

7- Lambert

8- Reap

از دیگر پژوهش‌ها تحلیل زمانی و مکانی توفان‌های تندری در دشت اردبیل به‌منظور مشخص کردن الگوهای سینوپتیک بود که با توجه به الگوها، بیشترین و کمترین درصد بارش رعد و برقی تعیین شد (نصیری قلعه‌بین و همکاران، ۱۳۹۸). تعیین توزیع زمانی و مکانی رعد و برق‌های ایران با استفاده از تصاویر ماهواره در دوره زمانی ۱۹۹۸ تا ۲۰۱۳ و با استفاده از تابع تراکم کرنل در نرم‌افزار جی.آی.اس.<sup>۱</sup> نشان‌دهنده بیشترین فراوانی رعد و برق در ماه می و آوریل و کمترین فراوانی در ماه‌های ژانویه و سپتامبر است (خورشیددوست و همکاران، ۱۳۹۶). از پژوهش‌های اخیر در هند می‌توان به پژوهش بالا<sup>۲</sup> و همکاران (۲۰۱۷) اشاره کرد که تکنیک‌های محاسباتی داده‌کاوی را برای پیش‌بینی رعد و برق و توفان‌های تندری بررسی کردند و به این نتیجه رسیدند که پیش‌بینی‌های بزرگ‌مقیاس به‌منظور کاهش خسارت ناشی از رعد و برق، کافی نیستند و توسعه پژوهش‌های کوچک‌مقیاس به‌دلیل اهمیت و کاربرد بیشتر ضروری است. در چین به‌منظور پیش‌بینی رعد و برق و توفان‌های تندری، در مقیاس کوچک و کوتاه‌مدت از چند مدل ماشینی داده‌کاوی استفاده کردند. نتایج به‌دست‌آمده گویای دقت و صحت بالاتر در پیش‌بینی با مدل ماشین بردار پشتیبان و مدل‌های مبتنی بر آن است (کیوو<sup>۳</sup> و همکاران، ۲۰۱۳). با توجه به اهمیت پیش‌بینی رعد و برق، هوو<sup>۴</sup> و همکاران (۲۰۱۴) مدل بهینه‌ای مبتنی بر الگوریتم سی‌فایو<sup>۵</sup> ایجاد کردند براساس نتایج حاصل‌شده، این مدل برای ارزیابی و پیش‌بینی رعد و برق نتایج مناسب و دقیقی را ارائه می‌دهد. بررسی‌های انجام‌شده نشان می‌دهد که درخت‌های تصمیم و خوشه‌بندی، الگوریتم‌های خوبی برای پیش‌بینی هستند و فن داده‌کاوی می‌تواند رفتارها و روندهای آینده را پیش‌بینی کند و تصمیم‌های پیشگیرانه و کاهش زمان برای حل مشکلات را ممکن می‌سازد (چاون و تاکور<sup>۶</sup>، ۲۰۱۴)؛ همچنین استفاده از تکنیک‌های هوش مصنوعی و داده‌کاوی می‌تواند به‌طور قابل توجهی مهارت پیش‌بینی را برای انواع مختلف شرایط آب‌وهوایی با تأثیر بالا بهبود بخشد (مک‌گاورن<sup>۷</sup> و همکاران، ۲۰۱۷).

در کانادا با استفاده از کوواریات<sup>۸</sup> جغرافیایی و زمانی، مشاهدات هواشناسی و مشاهدات رادیوسوند<sup>۹</sup> مدل‌های پیش‌بینی رعد و برق ۶ ساعته و ۲۴ ساعته‌ای از آوریل تا اکتبر برای پیش‌بینی رعد و برق منطقه البرتا در کانادا تهیه شد؛ براساس نتایج، پیش‌بینی با این مدل‌ها در مناطق کوه راکی و دامنه کوه دقت قابل قبول ۸۵ درصدی داشت و تکنیک داده‌کاوی و طبقه‌بندی جنگل تصادفی به‌مثابه روش مدل‌سازی مناسبی برای پیش‌بینی رعد و برق معرفی شد (بولین<sup>۱۰</sup> و همکاران، ۲۰۱۶). از پژوهش‌های اخیر پیش‌بینی وقوع رعد و برق با توجه به پارامترهای هواشناسی فشار، دما، رطوبت، سرعت باد و تکنیک‌های داده‌کاوی یادگیری ماشین است. نتایج ارزیابی نشان می‌دهد که مدل‌های ماشینی مهارت پیش‌بینی آماری قابل توجهی دارند و تکنیک‌های داده‌کاوی نیز برای مقایسه الگوهای توزیع داده، به‌صورت مکانی و زمانی در بین ایستگاه‌ها سودمند است (مستجایی<sup>۱۱</sup> و همکاران، ۲۰۱۹).

با گذشت زمان پژوهش‌های انجام‌شده در سطح جهان به‌منظور پیش‌بینی احتمال وقوع رعد و برق، همگی براساس استفاده و ارزیابی مدل‌های غیر هیبریدی و هیبریدی به‌منظور افزایش دقت در امر پیش‌بینی با

1- Geographic Information System (GIS)

2- Bala

3- Qiu

4- Hou

5- Induction of Decision Trees (C5)

6- Chauhan & Thakur

7- McGovern

8- Covariates

9- Radiosonde

10- Blouin

11- Mostajabi

در نظر گرفتن تأثیر پارامترهای اقلیمی و جغرافیایی در وقوع رعد و برق است؛ در حالی که پژوهش‌های داخلی تنها مبتنی بر توفان‌های تندری یا تعیین فراوانی یا پراکنش رعد و برق است و به پیش‌بینی احتمال وقوع رعد و برق و کاربرد تکنیک داده‌کاوی در پیش‌بینی توجّه چندانی نشده است. پیش‌بینی احتمال وقوع رعد و برق در ایران به دلیل حوادث خطرناک ناشی از آن نظیر سیلاب، توفان‌های تندری، بادهای شدید و... به‌ویژه در مناطق شمالی کشور مهم است. رشت، بزرگ‌ترین شهرستان برنج‌خیز کشور است و ۱۱٪ برنج مورد نیاز کشور در این استان تولید می‌شود؛ بنابراین اهمیت خسارت و آسیب به شالیزارها به‌علت وقوع رعد و برق، بارش شدید و وزش باد قابل چشم‌پوشی نیست. ورس (خوابیدگی ساقه برنج) و خطر آلودگی شالیزارها به بیماری بلاس، بسیار مهم است. بالآمدن آب رودخانه، آب‌گرفتگی معابر در اثر رعد و برق شدید و بارش باران، مشکلات تردد افراد و ترافیک، خسارت به منازل مسکونی، ساختمان‌ها و امکان ریزش پل نمونه‌های دیگری از آسیب‌ها هستند؛ از جمله گزارش‌های ثبت‌شده در استان رشت، خسارت ۱۶۰ میلیارد تومانی به تأسیسات و شبکه‌های برق در مهر ۱۳۹۴ به‌علت شرایط نامساعد جوی، رویداد رعد و برق، بارش شدید باران و جاری شدن سیل است؛ همچنین وجود تلفات جانی به‌علت برق‌گرفتگی با رعد و برق یا در اثر حوادث ناشی از آن نیز، مهم تلقی می‌شود؛ برای مثال گزارش ثبت‌شده در مهر ۱۳۹۱ نشان می‌دهد که ریزش شدید باران در رشت و مشکل اتّصال برق منجر به تلفات جانی شده است؛ بنابراین پیش‌بینی احتمال وقوع رعد و برق در آینده به‌منظور اقدام به‌موقع و کاهش خسارت می‌تواند مؤثر باشد.

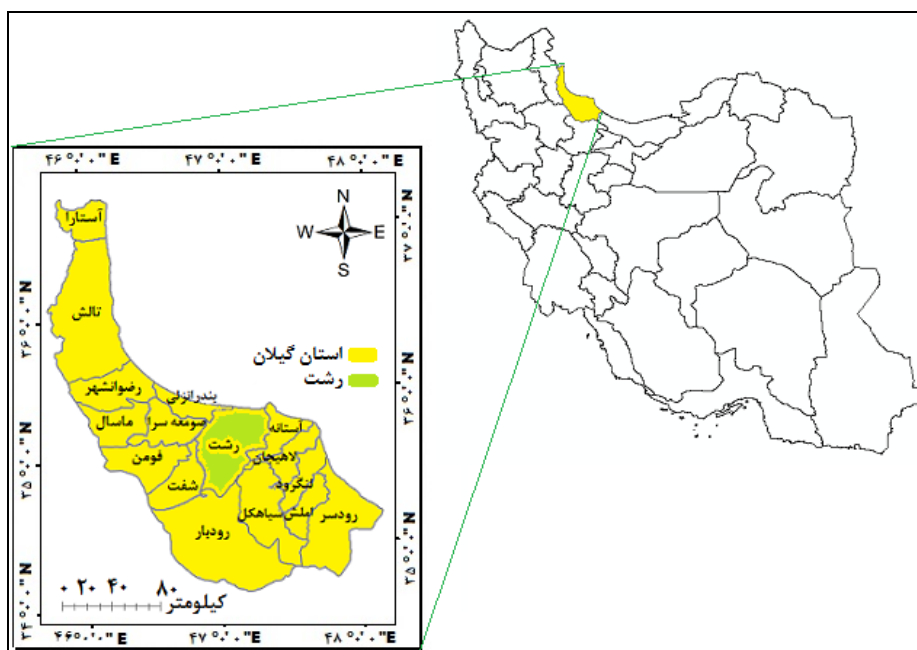
هدف اصلی نوشتار پیش رو استفاده از داده‌های زمینی ثبت‌شده از وقوع و عدم وقوع رعد و برق (داده‌های باینری) و تأثیر پارامترهای هواشناسی مرتبط (دما، رطوبت نسبی، ابرناکی، سرعت باد، جهت باد، فشار هوا و رعد و برق در روز گذشته) به‌منظور برآورد احتمال وقوع رعد و برق در روز آینده با استفاده از دانش داده‌کاوی (درخت‌ها و مدل‌های شبکه عصبی) و ارزیابی و تعیین مدل بهینه به‌منظور کاهش خسارت در آینده است.

## مواد و روش‌ها

شهر رشت مرکز استان گیلان و یکی از قطب‌های کشاورزی و گردشگری در کشور محسوب می‌شود. آب‌وهوای رشت معتدل کاسپین و شبه‌مدیترانه‌ای با زمستان‌های سرد و مرطوب و تابستان‌های گرم و شرجی است. جمعیت این شهر براساس آخرین سرشماری در آبان ۱۳۹۵، ۶۷۹۹۹۵ نفر با مساحت ۱۸۰ کیلومتر مربع است و کمابیش در مرکز جلگه‌ای گیلان بین "۴۵' ۳۵" ۴۹° طول شرقی و "۳۰' ۱۶" ۳۷° عرض شمالی قرار گرفته است (شکل ۱). در پژوهش حاضر به‌منظور پیش‌بینی رعد و برق از داده‌های جوی مربوط به سال‌های ۱۳۹۰ تا ۱۳۹۶ ایستگاه هواشناسی رشت استفاده شد. داده‌های مربوطه از سازمان هواشناسی گرفته شد. داده‌های ثبت‌شده رعد و برق از ایستگاه‌های زمینی به‌صورت داده‌های باینری (وجود یا عدم وجود رعد و برق) و سایر داده‌های مربوط به پارامترهای جوی (دما، رطوبت نسبی، ابرناکی، سرعت باد، جهت باد، فشار هوا و رعد و برق در روز گذشته) به‌صورت میانگین روزانه استفاده شد.

## نرمال‌سازی داده‌ها

از آنجاکه نرمال نبودن داده‌ها منجر به کاهش دقت و سرعت در مدل‌های پیش‌بینی خواهد شد (ژو و همکاران، ۲۰۰۷) پیش از ورود داده‌ها به مدل‌های استفاده‌شده، داده‌ها با استفاده از رابطه ۱ در دامنه صفر و یک نرمال شدند:



شکل ۱. موقعیت منطقه مورد بررسی

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

رابطه ۱

در رابطه بالا،  $x$ : داده نرمال شده،  $x_i$ : داده خام و  $X_{\max}$  و  $X_{\min}$ : داده حداکثر و حداقل هستند.

### متوازن سازی داده‌ها

به دلیل باینری بودن داده‌های مربوط به رعد و برق و تفاوت این داده‌ها در مقایسه با پارامترهای جوی که به صورت میانگین روزانه استفاده شدند، نوعی نامتوازنی در کلاس‌های داده‌ها در مدل‌های پیش‌بینی، به‌ویژه درخت‌های کارت، چاید<sup>۱</sup> و سی فایو مشاهده شد؛ در واقع در طبقه‌بندی داده‌های نامتوازن ممکن است طبقه‌ای که حداقل تعداد داده را دارد، صحت کمتری داشته باشد (جین<sup>۲</sup> و همکاران، ۲۰۱۶) که این امر پیش‌بینی را با مشکل روبه‌رو خواهد کرد. به منظور متوازن سازی داده‌های رعد و برق و حل مشکل عدم تعادل داده‌ها و طبقه‌های داده از رویکردهای سطح داده با استفاده از الگوریتم‌های روس<sup>۳</sup> و راس<sup>۴</sup> در نرم‌افزار ریپد ماینر<sup>۵</sup> استفاده شد. عملکرد این الگوریتم‌ها افزایش نمونه‌های مربوط به طبقه‌های حداقل و کاهش نمونه‌های مربوط به طبقه‌های حداکثر است (ژانگ<sup>۶</sup> و همکاران، ۲۰۱۶)؛ سپس روند تغییرات متغیرهای مستقل و وابسته با تعیین خصوصیات آماری و همبستگی پیرسون<sup>۷</sup> بین متغیرها با نرم‌افزار اس.پی.اس.اس<sup>۸</sup> به منظور کاهش خطای مدل‌ها در پیش‌بینی انجام شد.

### داده‌کاوی

پس از نرمال و متوازن سازی داده‌ها، داده‌کاوی صورت گرفت. در بیشتر اوقات، پیش از اجرای الگوریتم‌های

1- CHi-squared Automatic Interaction Detector (CHAID)

2- Jian

3- Random Over Sampling (ROS)

4- Random Under Sampling (RUS)

5- Rapid Miner

6- Zhang

7- Peirson

8- Statistical Package for Social Science (SPSS)

داده‌کاوی، لازم است که برحسب نیاز پیش‌پردازش‌هایی مانند جمع‌آوری داده‌های مرتبط، یکسان کردن منابع داده‌ها، پاکسازی داده‌ها، تبدیل مقادیر پیوسته به گسسته (در صورت نیاز)، انتخاب ویژگی‌های مناسب، کاهش حجم داده‌ها (برای مثال با تکنیک‌های کاهش ابعاد) و... بر روی داده‌ها انجام شود (هان<sup>۱</sup> و همکاران، ۲۰۰۶: ۳۶۰). بدین منظور پس از پیش‌پردازش داده‌ها، از نرم‌افزار اس.پی.اس.اس. مودلر<sup>۲</sup> برای پیش‌بینی احتمال وقوع و عدم وقوع رعد و برق در آینده با استفاده از درخت‌های کارت، چاید و سی فایو و مدل‌های شبکه عصبی پرسپترون چندلایه<sup>۳</sup> و تابع پایه‌ای شعاعی<sup>۴</sup> و ماشین بردار پشتیبان<sup>۵</sup> استفاده شد. در این مدل‌ها متغیر وابسته، وقوع و عدم وقوع رعد و برق و متغیرهای مستقل عوامل مؤثر بر رعد و برق شامل دما، رطوبت نسبی، ابرناکی، سرعت باد، جهت باد، فشار هوا و رعد و برق در روز گذشته است و به منظور عملکرد بهینه مدل، ۷۰٪ داده‌ها برای مرحله آموزش و ۳۰٪ برای مرحله آزمون مدل استفاده شده است. نتایج حاصل از درخت‌ها به صورت عدد و قانون و نتایج حاصل از شبکه‌های عصبی به صورت عددی با استفاده از آماره‌ها ارزیابی شد.

### ارزیابی کارایی مدل

نتایج حاصل از خروجی مدل‌ها با استفاده از روابط ۲ تا ۹ بررسی شد. به منظور ارزیابی کارایی مدل‌ها از ماتریس درهم‌ریختگی و معیارهای مقایسه‌ای صحت (رابطه ۲)، دقت (رابطه ۳)، حساسیت (رابطه ۴)، میانگین هارمونیک (رابطه ۵)، تشخیص (رابطه ۶)، نرخ مثبت کاذب (رابطه ۷)، نرخ منفی کاذب (رابطه ۸) و ریشه میانگین مربعات خطا<sup>۹</sup> (رابطه ۹) استفاده شد. ماتریس درهم‌ریختگی مشخص‌کننده رده‌بندی‌های واقعی و پیش‌بینی‌شده با مدل است و اطلاعات ارزشمندی دارد (کوهاوی<sup>۷</sup>، ۱۹۹۸). منحنی راک<sup>۸</sup> یکی از روش‌های شاخص برای مشخص کردن قدرت مدل‌های پیش‌بینی است. سطح زیر این منحنی (آ.ی.وسی.<sup>۸</sup>) دو مقدار ۰/۵ تا ۱ دارد که مشخص می‌کند مدل با چه احتمالی موقعیت مثبت را نسبت به منفی انتخاب می‌کند. رابطه‌های ۲ تا ۹ در ارزیابی کارایی مدل مورد توجه است.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

رابطه ۲

$$\text{Precision} = \frac{TP}{TP+FP}$$

رابطه ۳

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

رابطه ۴

$$\text{Harmonic Mean} = \frac{2 * P * S}{P + S}$$

رابطه ۵

$$\text{Specificity} = \frac{TN}{TN+FP}$$

رابطه ۶

$$\text{نرخ مثبت کاذب} = \frac{FP}{FP+TN}$$

رابطه ۷

- 1- Han
- 2- Spss Modeler
- 3- Multi Layer Perceptron (MLP)
- 4- Radial Basis Function (RBF)
- 5- Support Vector Machine (SVM)
- 6- Root Mean Square Error (RMSE)
- 7- Kohavi
- 8- Receiver Operating Characteristic (ROC)
- 9- Area Under Curve (AUC)

$$\text{نسبت منفی کاذب} = FN / (FN + TP)$$

رابطه ۸

$$RMSE = \sqrt{1/N \sum (P-O)^2}$$

رابطه ۹

منظور از TN: نرخ منفی صحیح؛ FP: نرخ مثبت کاذب؛ FN: نرخ منفی کاذب؛ TP: نرخ مثبت صحیح؛ O: مقدار مشاهده‌شده؛ P: مقدار پیش‌بینی شده و N: تعداد داده‌ها است.

### نتایج

#### خصوصیات آماری متغیرها

پس از پیش‌پردازش داده‌ها، خصوصیات آماری متغیرها طی هفت سال نشان می‌دهد که میانگین سالانه دما و رطوبت نسبی از سال ۱۳۹۰ تا ۱۳۹۶ متغیر و به مقدار کم کاهش یافته است و تغییرات کم انحراف معیار، نوسانات کم دما و رطوبت را نشان می‌دهد. میانگین جهت باد و ابرناکی متغیر است و مقادیر متفاوت انحراف معیار نشان می‌دهد که نوسانات زیادی در میزان جهت باد و ابرناکی در منطقه وجود دارد و میزان انحراف از میانگین و تغییرات متوسط سالانه سرعت باد و فشار هوا اندک است؛ همچنین طی هفت سال مقادیر بالای انحراف معیار گویای نوسانات بسیار رعد و برق و میانگین متوسط رعد و برق نشان‌دهنده افزایش تدریجی رعد و برق و کاهش آن در سال ۱۳۹۶ است (جدول ۱).

#### نتایج حاصل از همبستگی بین متغیرهای مستقل با رعد و برق شهر رشت

نتایج حاصل از همبستگی پیرسون نشان می‌دهد بیشترین میزان همبستگی رعد و برق با پارامتر رعد و برق در روز گذشته و حداقل همبستگی با متغیر سرعت باد است (جدول ۲).

جدول ۱. نتایج حاصل از خصوصیات آماری متغیرهای مستقل و وابسته

پارامترهای هواشناسی	سال ۱۳۹۰		سال ۱۳۹۱		سال ۱۳۹۲		سال ۱۳۹۳		سال ۱۳۹۴		سال ۱۳۹۵		سال ۱۳۹۶	
	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل
دما	۰/۹۶	۰/۳۵	۰/۹۱	۰/۳۰	۰/۸۶	۰/۴۵	۰/۹۲	۰/۲۷	۰/۹۴	۰/۲۴	۰/۸۹	۰/۱۶	۰/۹۶	۰/۳۰
رطوبت نسبی	۰/۷۰	۰/۱۶	۰/۷۴	۰/۱۱	۰/۶۹	۰/۱۲	۰/۷۲	۰/۱۷	۰/۶۷	۰/۱۹	۰/۶۹	۰/۱۷	۰/۶۷	۰/۱۸
جهت باد	۰/۹۷	۰/۵۱	۰/۹۷	۰/۴۴	۰/۹۱	۰/۴۴	۰/۹۸	۰/۳۸	۰/۹۴	۰/۴۴	۰/۹۸	۰/۳۱	۰/۹۷	۰/۰۰۷
ابرنایی	۰/۶۷	۰/۲۴	۰/۵۶	۰/۲۶	۰/۶۶	۰/۲۲	۰/۵۷	۰/۲۵	۰/۷۰	۰/۲۵	۰/۶۸	۰/۲۶	۰/۶۴	۰/۲۴
سرعت باد	۰/۴۷	۰/۰۴	۰/۳۷	۰	۰/۳۳	۰	۰/۳۸	۰/۰۶	۱	۰/۰۴	۰/۶۳	۰/۰۲	۰/۶۶	۰/۰۹
فشار هوا	۰/۶۸	۰/۹۸	۰/۸۷	۰/۰۷	۰/۷۱	۰/۰۸	۰/۵۹	۰/۰۸	۰/۸۳	۰/۱۱	۰/۶۶	۰/۱۷	۰/۵۲	۰/۱۳
رعد و برق روز گذشته	۰/۶۱	۰/۴۹	۰/۶۳	۰/۴۸	۰/۶۴	۰/۴۸	۰/۶۰	۰/۴۹	۰/۶۱	۰/۴۹	۰/۶۲	۰/۴۸	۰/۵۴	۰/۵۰
رعد و برق	۰/۶۱	۰/۴۹	۰/۶۳	۰/۴۸	۰/۶۴	۰/۴۸	۰/۶۰	۰/۴۹	۰/۶۱	۰/۴۹	۰/۶۲	۰/۴۸	۰/۵۴	۰/۵۰

جدول ۲. نتایج حاصل از همبستگی متغیرهای مستقل با رعد و برق شهر رشت

رعد و برق	رعد و برق روز گذشته	دما	فشار هوا	رطوبت نسبی	ابرنایی	جهت باد	سرعت باد
۰/۱۷۳	۰/۵۲۷	۰/۱۳۷	-۰/۱۲۷	۰/۱۲۵	-۰/۰۴۵	۰/۰۱۴	



### پیش‌بینی درصد وقوع و عدم وقوع رعد و برق با درخت‌ها

با استفاده از الگوریتم‌های کارت، چاید و سی فایو، وقوع و عدم وقوع رعد و برق با توجه به داده‌های مشاهده‌شده و براساس قوانین ارائه‌شده، پیش‌بینی شد. براساس درخت کارت، میزان وقوع رعد و برق ۷۷٪ است. درخت چاید میزان وقوع رعد و برق در روزهای آبی را ۵۴٪ و میزان عدم وقوع آن را ۴۶٪ پیش‌بینی کرد و درخت سی فایو احتمال وقوع را براساس قوانین ذکرشده ۷۵٪ پیش‌بینی می‌کند. با توجه به نتایج به‌دست‌آمده، احتمال وقوع رعد و برق در روز آبی بیشتر از عدم وقوع آن است. در قوانین ذکرشده کمیت یک، نشان‌دهنده احتمال وقوع رعد و برق و کمیت صفر، نشانه عدم وقوع رعد و برق است (جدول ۳).

### پیش‌بینی درصد وقوع و عدم وقوع رعد و برق با مدل‌های شبکه عصبی مصنوعی

مدل‌های پرسپترون چندلایه و تابع پایه‌ای شعاعی احتمال وقوع رعد و برق را به ترتیب ۵۷٪ و ۶۴٪ پیش‌بینی کردند و مدل ماشین بردار پشتیبان میزان وقوع رعد و برق در روز آینده را ۶۰٪ و عدم وقوع آن را ۴۰٪ پیش‌بینی کرده است که نتایج نشان‌دهنده احتمال وقوع رعد و برق در روزهای آبی است و با نتایج حاصل از درخت‌ها در وقوع احتمالی رعد و برق همخوانی دارد (جدول ۴).

نتایج پیش‌بینی حاصل از درخت‌ها و مدل‌های شبکه عصبی مصنوعی به‌وضوح نشان می‌دهد، با توجه به مقادیر مشاهده‌شده از داده‌های مربوط به احتمال وقوع یا عدم وقوع رعد و برق از میان مدل‌های مورد استفاده در پیش‌بینی، درخت چاید و مدل ماشین بردار پشتیبان نزدیک‌ترین شباهت پیش‌بینی به مقادیر مشاهده‌شده در احتمال وقوع رعد و برق در آینده را دارند (شکل ۲).

### پیش‌بینی احتمال وقوع رعد و برق در ماه‌های مختلف در آینده

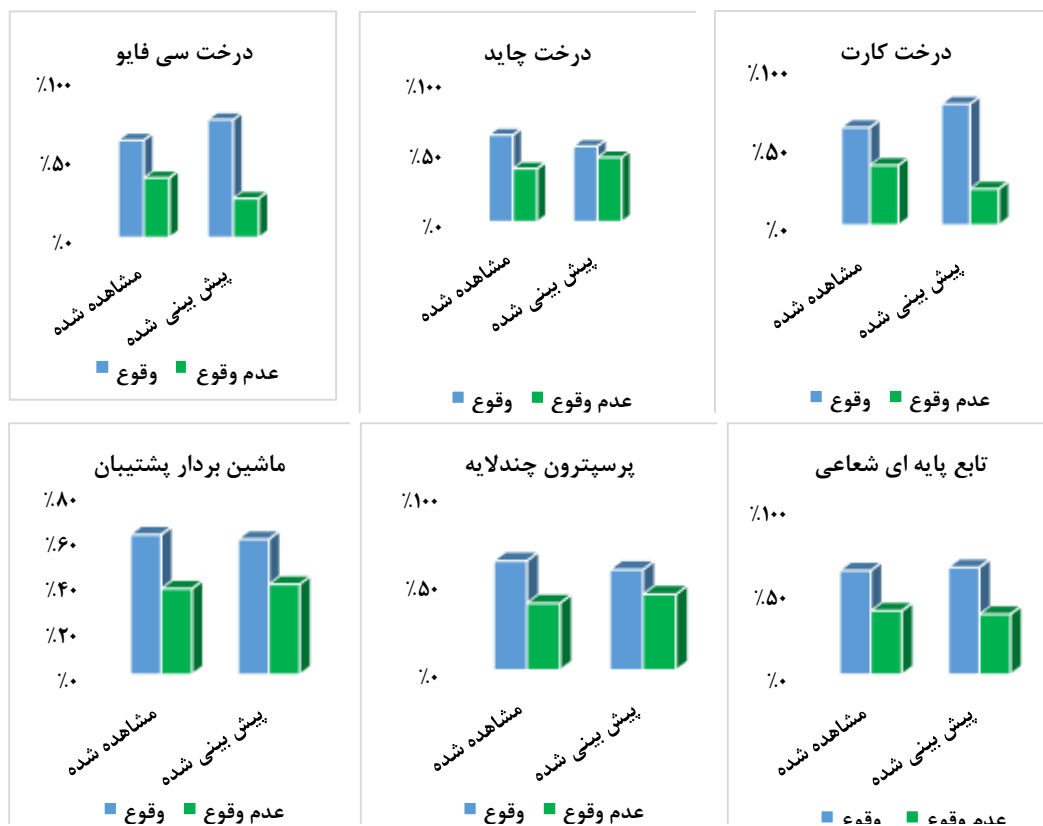
براساس نتایج به‌دست‌آمده از خروجی مدل‌ها، احتمال وقوع رعد و برق در فصل بهار بیشتر از سایر فصل‌ها است و از فصل بهار تا زمستان وقوع رعد و برق روند کاهشی دارد و در فصل زمستان احتمال وقوع بسیار کم و ناچیز است. در راستای پیش‌بینی با درخت‌ها و مدل‌های شبکه عصبی، احتمال وقوع رعد و برق در آینده در ماه‌های اردیبهشت، خرداد و تیر نسبت به سایر ماه‌ها حداکثر است. مدل‌ها روند یکسانی را برای وقوع رعد و برق در ماه فروردین پیش‌بینی کردند و احتمال رویداد رعد و برق در مهرماه نسبت به سایر ماه‌های پاییز بیشتر پیش‌بینی شده است. احتمال وقوع در زمستان دارای حداقل مقدار و عدم وقوع است و پیش‌بینی انجام‌شده، سیر نزولی رعد و برق از فصل بهار تا زمستان را نشان می‌دهد (جدول ۵).

جدول ۳. پیش‌بینی احتمال وقوع و عدم وقوع رعد و برق در روز آینده با درخت

درخت‌ها	وقوع	عدم وقوع	قوانین
کارت	۷۷٪	۲۳٪	رعد و برق در روز گذشته $\geq 0/5$ و $0/715 < \text{دما}$ $\leftarrow 1$ $0/715 \geq \text{دما} \leftarrow 0$
چاید	۵۴٪	۴۶٪	فشار $\geq 0/378$ و فشار $< 0/378$ و فشار $\geq 0/5$ و فشار $< 0/5$ و دما $< 0/844$ $\leftarrow 1$ رطوبت نسبی $< 0/779$ و رطوبت $\geq 0/824$ و رطوبت $< 0/779$ و دما $\geq 0/656$ و رطوبت $< 0/824$ $\leftarrow 0$
سی فایو	۷۵٪	۲۵٪	رعد و برق در روز گذشته $\geq 0$ و ابرناکی $< 0/224$ $\leftarrow 1$ $0/715 \geq \text{دما}$ و ابرناکی $\geq 0/224$ $\leftarrow 0$

جدول ۴. پیش‌بینی احتمال وقوع و عدم وقوع رعد و برق در روز آینده با شبکه عصبی مصنوعی

مدل‌ها	وقوع	عدم وقوع
تابع پایه‌ای شعاعی	۶۴٪	۳۶٪
پرسپترون چندلایه	۵۷٪	۴۳٪
ماشین بردار پشتیبان	۶۰٪	۴۰٪



شکل ۲. نتایج حاصل از پیش‌بینی درخت‌ها و مدل‌های شبکه‌ی عصبی مصنوعی

جدول ۵. پیش‌بینی رخداد آینده رعد و برق در ماه‌های مختلف

زمستان	پاییز	تابستان	بهار	احتمال وقوع رعد و برق	
۲٪	۱۰٪	۲۹٪	۳۶٪	۷۷٪	کارت
۲٪	۸٪	۲۱٪	۲۳٪	۵۴٪	چاید
۲٪	۱۰٪	۲۷٪	۳۶٪	۷۵٪	سی فایو
۲٪	۸٪	۲۴٪	۳۱٪	۶۴٪	پایه‌ای شعاعی
۲٪	۸٪	۲۱٪	۲۶٪	۵۷٪	پرسپترون چندلایه
۲٪	۹٪	۲۱٪	۲۷٪	۶۰٪	ماشین بردار پشتیبان

زمستان			پاییز			تابستان			بهار		فصل‌ها
اسفند	بهمن	دی	آذر	آبان	مهر	شهریور	مرداد	تیر	خرداد	اردیبهشت	فروردین
۰.۰۷۶٪	۰.۰۷۶٪	۰.۰۲۵٪	۰.۰۷۶٪	۰.۳٪	۰.۷٪	۰.۷٪	۰.۷٪	۰.۱۶٪	۰.۱۸٪	۰.۱۳٪	۰.۵٪
۰.۰۷۶٪	۰.۰۷۶٪	۰.۰۲۵٪	۰.۰۷۶٪	۰.۳٪	۰.۵٪	۰.۵٪	۰.۵٪	۰.۱۱٪	۰.۹٪	۰.۹٪	۰.۵٪
۰.۰۷۶٪	۰.۰۷۶٪	۰.۰۲۵٪	۰.۰۷۶٪	۰.۲٪	۰.۷٪	۰.۶٪	۰.۶٪	۰.۱۵٪	۰.۱۷٪	۰.۱۳٪	۰.۵٪
۰.۰۷۶٪	۰.۰۷۶٪	۰.۰۲۵٪	۰.۱٪	۰.۲٪	۰.۵٪	۰.۵٪	۰.۶٪	۰.۱۳٪	۰.۱۳٪	۰.۱۲٪	۰.۵٪
۰.۰۱۵٪	۰.۱٪	۰.۰۱۵٪	۰.۰۷۶٪	۰.۳٪	۰.۵٪	۰.۴٪	۰.۵٪	۰.۱۱٪	۰.۱۱٪	۰.۱۱٪	۰.۵٪
۰.۰۱۵٪	۰.۰۷۶٪	۰.۰۱۵٪	۰.۱٪	۰.۲٪	۰.۶٪	۰.۵٪	۰.۵٪	۰.۱۱٪	۰.۱۲٪	۰.۱۲٪	۰.۴٪

### ارزیابی عملکرد درخت‌ها و مدل‌های شبکه‌ی عصبی مصنوعی با شاخص‌های یادشده

مدل بهینه از میان درخت‌ها و مدل‌های شبکه‌ی عصبی مصنوعی با استفاده از ماتریس درهم‌ریختگی و شاخص‌های دقت، صحت، حساسیت، تشخیص، میانگین هارمونیک، ریشه میانگین مربع خطا، نرخ مثبت و منفی کاذب ارزیابی شد. براساس نتایج به‌دست‌آمده، از میان درخت‌های پیش‌بینی، درخت چاید با مقادیر شاخص‌های صحت = ۰/۷۶۵،

دقت =  $0/853$  و تشخیص =  $0/794$  و مناسب بودن سایر شاخص‌های ریشه میانگین مربع خطا =  $0/483$  و نرخ مثبت کاذب =  $0/205$  در مقایسه با درخت‌های کارت و سی فایو بهترین عملکرد را دارد. درخت کارت از نظر شاخص حساسیت و نرخ منفی کاذب و درخت سی فایو از نظر پارامتر میانگین هارمونیک عملکرد مطلوبی دارند. از میان مدل‌های شبکه عصبی مصنوعی، مدل تابع پایه‌ای شعاعی از نظر حساسیت و نرخ منفی کاذب مقادیر مطلوب و یکسانی با مدل ماشین بردار پشتیبان دارد و مدل پرسپترون چندلایه، مقادیر مطلوبی در شاخص‌های تشخیص و نرخ مثبت کاذب دارد. مدل پرسپترون چندلایه با حداکثر مقادیر صحت =  $0/773$ ، دقت =  $0/855$ ، حساسیت =  $0/801$ ، میانگین هارمونیک =  $0/813$  و حداقل مقدار ریشه میانگین مربع خطا =  $0/475$  و نرخ منفی کاذب =  $0/198$ ، عملکرد بسیار مطلوبی در پیش‌بینی آینده رعد و برق دارد.

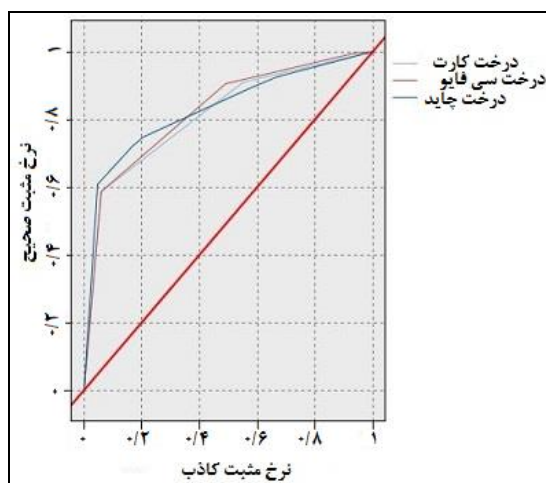
به‌طور کلی مدل پرسپترون چندلایه حداکثر صحت و دقت و کمترین مقدار ریشه میانگین مربعات خطا و درخت چاید بالاترین میزان تشخیص و حداقل مقدار برای نرخ مثبت کاذب را دارند (جدول ۶).

### تحلیل نتایج حاصل از درخت‌ها و مدل‌های شبکه عصبی مصنوعی با منحنی راک

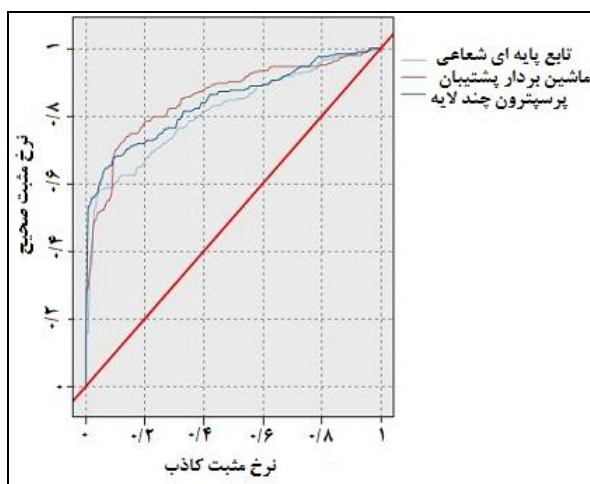
منحنی راک امکان مقایسه تصویری مجموعه‌ای از دسته‌بندی‌ها را با مدل فراهم می‌کند. مساحت زیر این منحنی هر چه به یک نزدیک‌تر باشد، عملکرد و قدرت مدل در پیش‌بینی بالاتر خواهد بود. براساس شکل ۲ و ۳ و جدول ۷، درخت چاید با مساحت زیر نمودار  $0/829$  و مدل ماشین بردار پشتیبان با مقدار  $0/853$  نسبت به سایر مدل‌ها برتری دارند و طبقه‌بندی بهتری را ارائه می‌دهند.

جدول ۶. نتایج حاصل از عملکرد درخت‌ها و مدل‌های شبکه عصبی مصنوعی با استفاده از شاخص‌ها

مدل‌ها	نرخ منفی صحیح	نرخ مثبت کاذب	نرخ مثبت کاذب	نرخ مثبت صحیح	صحت	دقت	حساسیت	تشخیص	میانگین هارمونیک	ریشه میانگین مربعات خطا	نرخ مثبت کاذب	نرخ منفی کاذب
کارت	۶۸	۸۳	۲۲	۲۲۰	۰/۷۳۲	۰/۷۲۶	۰/۹۰۹	۰/۴۵۰	۰/۸۰۷	۰/۵۷۶	۰/۵۴۹	۰/۰۹۰
چاید	۱۲۰	۳۱	۶۱	۱۸۱	۰/۷۶۵	۰/۸۵۳	۰/۷۴۷	۰/۷۹۴	۰/۷۹۷	۰/۴۸۳	۰/۲۰۵	۰/۲۵۲
سی فایو	۷۷	۷۴	۲۳	۲۱۹	۰/۷۵۳	۰/۷۴۷	۰/۹۰۴	۰/۵۰۹	۰/۸۱۸	۰/۴۹۶	۰/۴۹۰	۰/۰۹۵
پایه‌ای شعاعی	۹۲	۵۹	۴۸	۱۹۴	۰/۷۲۷	۰/۷۶۶	۰/۸۰۱	۰/۶۰۹	۰/۷۸۳	۰/۵۲۱	۰/۳۹۰	۰/۱۹۸
پرسپترون چندلایه	۱۱۱	۴۰	۵۹	۱۸۳	۰/۷۴۸	۰/۸۲۰	۰/۷۵۶	۰/۷۲۵	۰/۷۸۷	۰/۵۰۱	۰/۲۶۴	۰/۲۴۳
ماشین بردار پشتیبان	۱۱۰	۴۱	۴۸	۱۹۴	۰/۷۷۳	۰/۸۵۵	۰/۸۰۱	۰/۷۲۸	۰/۸۱۳	۰/۴۷۵	۰/۲۷۱	۰/۱۹۸



شکل ۳. تحلیل نتایج حاصل از درخت‌های کارت، چاید و سی فایو



شکل ۴. تحلیل نتایج حاصل از مدل‌های پایه‌ای شعاعی، پرسپترون چندلایه و ماشین بردار پشتیبان

جدول ۷. نتایج حاصل از درخت‌ها و مدل‌های شبکه عصبی مصنوعی با منحنی راک

منحنی راک	کارت	چاید	سی فایو	پایه‌ای شعاعی	پرسپترون چند لایه	ماشین بردار پشتیبان
سطح زیر نمودار راک	۰/۸۱۴	۰/۸۲۹	۰/۸۲۶	۰/۸۰۹	۰/۸۴۰	۰/۸۵۳

## بحث

رعد و برق از مهم‌ترین مخاطرات محیطی با خسارات فراوان است. از آنجاکه رعد و برق و حوادث ناشی از آن منجر به خسارات بسیاری می‌شوند، پیش‌بینی احتمال وقوع یا عدم وقوع این پدیده و تعیین روند وقوع آن در فصل‌های مختلف با دقت و صحت حداکثر و حداقل خطا بسیار مهم است. براساس نتایج به‌دست‌آمده از پژوهش حاضر احتمال وقوع رعد و برق در فصل بهار (اردیبهشت و خرداد) و تابستان (تیر) حداکثر است و در فصل زمستان به حداقل می‌رسد و روند نزولی دارد و احتمال وقوع رعد و برق در آینده بیش از عدم وقوع آن است؛ در واقع عامل وقوع توفان‌های تندری در فصل بهار و تابستان در بررسی همدیدی توفان‌ها در آستارا عوامل توپوگرافیکی (دریا، رطوبت دریا، ارتفاع از دریا) منطقه است (مفتاحی نمین و صلاحی، ۱۳۹۲).

بررسی انجام‌شده از رخداد توفان‌های تندری بهاره در ایران در بازه زمانی ۵۱ ساله از ۱۹۶۰ تا ۲۰۱۰ نشان می‌دهد که شهر رشت از نظر میانگین و فراوانی توفان‌های تندری بهاره از بین پنج گروه در رتبه سوم قرار دارد و حداکثر بارش‌های رعد و برقی در ماه اردیبهشت رخ می‌دهد (قویدل رحیمی و همکاران، ۱۳۹۳)؛ همچنین حداکثر فراوانی وقوع توفان‌های تندری با استفاده از داده ۴۶ ایستگاه سینوپتیک کشور در دوره زمانی ۱۳۶۵ تا ۱۳۸۸ در فصل بهار اتفاق می‌افتد (خالصی، ۱۳۹۳). به‌طور کلی در پیش‌بینی رعد و برق و توفان‌های تندری پیش‌بینی‌های کوچک‌مقیاس به‌دلیل خطای کمتر کاربردی و مؤثرتر از پیش‌بینی‌های بزرگ‌مقیاس هستند (بالا و همکاران، ۲۰۱۷)؛ برای مثال بررسی توفان‌های تندری و بارش‌های ناشی از آن در اهر نشان داد که بیشتر بارش‌ها در فصل بهار و تابستان رخ می‌دهد (جلالی و همکاران، ۱۳۸۵)؛ همچنین بیشترین رخداد وقوع رعد و برق و توفان‌های تندری در تبریز در فصل بهار و ماه می روی می‌دهد (خوشحال دستجردی و قویدل رحیمی، ۱۳۸۶) و تحلیل ارتباط بین فراوانی وقوع توفان‌های تندری و روند دما در مشهد مشخص کرد که میانگین دما و وقوع توفان‌های تندری رابطه مستقیمی دارند و حداکثر میانگین دما در ماه جولای و حداکثر توفان‌های تندری در فصل بهار رخ داده است (موسوی و همکاران، ۱۳۹۶).

پژوهش‌های انجام‌شده گویای اهمیت و وقوع پدیده رعد و برق و بارش‌های رعد و برقی در فصل بهار به‌ویژه ماه

اردیبهشت هستند که مبین احتمال بسیار زیاد وقوع این پدیده در فصل بهار است. پژوهشگران از روش‌های متفاوتی به منظور پیش‌بینی رعد و برق استفاده می‌کنند؛ برای مثال باساک<sup>۱</sup> و همکاران (۲۰۱۲) در کلکته روزهای دارای رعد و برق را با مدل‌های مبتنی بر شبکه عصبی برای سه سال آینده پیش‌بینی کردند؛ افزون بر این روش‌ها، تکنیک داده‌کاوی روش مناسبی برای پیش‌بینی است؛ در واقع داده‌کاوی، مرحله‌ای از فرایند استخراج دانش است که سه فاز را شامل می‌شود: پیش‌پردازش داده‌ها، داده‌کاوی و پردازش دانش حاصل شده، نحوه پیش‌پردازش داده‌ها و متغیرهای منتخب، تأثیر قابل توجهی در کشف دانش دارد. نتایج به‌دست‌آمده در پژوهش حاضر نشان می‌دهد که پیش‌بینی با استفاده از تکنیک داده‌کاوی امکان‌پذیر و مؤثر است؛ در این راستا پژوهش‌های بولین و همکاران (۲۰۱۶) و مک‌گاورن و همکاران (۲۰۱۷) گویای قابلیت بالای تکنیک‌های داده‌کاوی و هوش مصنوعی در پیش‌بینی رعد و برق است. در کانادا با استفاده از مشاهدات شبکه ردیابی رعد و برق و داده‌های به‌دست‌آمده از مرکز آب‌وهوایی طی سال‌های ۲۰۰۰ تا ۲۰۰۱ مدل آماری مبتنی بر درخت کارت را روش مناسبی برای پیش‌بینی احتمال وقوع رعد و برق تشخیص دادند (باروس<sup>۲</sup> و همکاران، ۲۰۰۵)؛ همچنین مدل غیر خطی پرسپترون چندلایه نسبت به مدل خطی در پیش‌بینی رعد و برق و بادهای شدید ناشی از آن عملکرد بهتر و دقیق‌تری را نشان داده است (چادوری و میدی<sup>۳</sup>، ۲۰۱۳). براساس پژوهش چاوون و تاکور (۲۰۱۴) تکنیک داده‌کاوی می‌تواند رفتارها و روندهای آینده را پیش‌بینی کند و تصمیم‌های پیشگیرانه و کاهش زمان برای حل مشکلات و کاهش خسارت را ممکن می‌سازد.

از آنجاکه افزایش دقت مدل، رابطه مستقیمی با افزایش میزان و تأثیر متغیرهای ورودی به مدل را دارد (راجیوان و همکاران، ۲۰۱۲). به‌منظور افزایش صحت مدل‌ها پس از نرمال و متوازن سازی داده‌های مؤثر در وقوع رعد و برق، داده‌ها با استفاده از الگوریتم‌های گوناگون آموزش داده شدند. براساس نتایج به‌دست‌آمده از سه درخت کارت، چاید و سی فایو درخت کارت و سی فایو از نظر شاخص‌های محدودی مطلوبیت داشتند و فاقد حداکثر صحت و دقت در پیش‌بینی رعد و برق در روز آتی بودند؛ در حالی که درخت چاید در ۰/۷۶ موارد، پیش‌بینی صحیحی با دقت ۰/۸۵ داشته است و درصد وقوع رعد و برق را ۰/۵۴ پیش‌بینی کرده است که با مقدار واقعی ۰/۶۲ بسیار شباهت دارد و از میان مدل‌های شبکه عصبی مصنوعی، مدل ماشین بردار پشتیبان با حداکثر مطلوبیت با پیش‌بینی ۰/۷۷ صحت و دقت ۰/۸۵ و پیش‌بینی ۰/۶۰، احتمال وقوع رعد و برق نسبت به مدل‌های تابع پایه‌ای شعاعی و پرسپترون چندلایه اولویت و برتری دارد. از نظر طبقه‌بندی و سطح زیر نمودار راک از میان درخت‌ها، درخت چاید با مقدار ۰/۸۲۹ و مدل ماشین بردار پشتیبان با مقدار ۰/۸۵۳ برتری دارد. نتایج عددی حاصل شده و مشابهت این پیش‌بینی با مقادیر واقعی نشان می‌دهد که درخت‌ها و شبکه عصبی مصنوعی در پیش‌بینی احتمال وقوع رعد و برق در آینده مؤثر هستند. در پیش‌بینی رعد و برق و توفان‌های تندری مدل ماشین بردار پشتیبان و مدل‌های مبتنی بر این مدل، به‌دلیل ساختار غیر خطی و کاهش میزان خطا در مرحله آموزش داده‌ها، حداکثر صحت و دقت را در پیش‌بینی دارند (کیو و همکاران، ۲۰۱۳). براساس نتایج به‌دست‌آمده در پیش‌بینی رعد و برق از میان مدل‌های پیش‌بینی‌کننده، درخت چاید و مدل ماشین بردار پشتیبان نسبت به سایر مدل‌ها عملکرد بهینه‌ای دارند و قابلیت پیش‌بینی بهتری را ارائه می‌دهند.

1- Basak

2- Burrows

3- Chaudhuri &amp; Middey

### نتیجه‌گیری

با توجه به نتایج به‌دست‌آمده از خروجی مدل‌ها، مشخص شد که احتمال وقوع رعد و برق در شهر رشت بسیار است. مدل‌ها احتمال وقوع رعد و برق در ماه فروردین را با روند یکسانی نشان می‌دهند؛ اما حداکثر وقوع رعد و برق در بهار (اردیبهشت و خرداد) به دلیل شرایط ناپایدار جو و تابستان (تیر) بیشتر از پاییز و زمستان است و از بهار تا زمستان سیر نزولی دارد و در زمستان به حداقل می‌رسد. بررسی میان مدل‌های پیش‌بینی‌کننده درخت و شبکه عصبی نشان داد که درخت چاید با میزان تشخیص ۰/۷۹۴ و حداقل میزان نرخ مثبت کاذب ۰/۲۰۵ و مدل ماشین بردار پشتیبان با پیش‌بینی صحیح ۰/۷۷۳ و نرخ خطای ۰/۴۷۵ و دقت ۰/۸۵۵ نسبت به سایر مدل‌ها به دلیل برتری آماره‌های یادشده عملکرد مطلوبتری دارند و به‌طور کلی از ارزیابی درخت و مدل شبکه عصبی بهینه، مدل ماشین بردار پشتیبان با مقداری تفاوت جزئی در مطلوبیت شاخص‌های صحت = ۰/۷۷۳، دقت = ۰/۸۵۵، میانگین هارمونیک = ۰/۸۱۳، ریشه میانگین مربع خطا = ۰/۴۷۵ و نرخ منفی کاذب = ۰/۱۹۸ به‌مثابه مدل بهینه در پیش‌بینی رعد و برق در آینده مشخص شد و با توجه به خروجی‌های قابل اعتماد با حداکثر صحت، دقت و کمترین میزان خطا در پیش‌بینی، مدل ماشین بردار پشتیبان کارایی مناسب را دارد و می‌تواند برای پیش‌بینی احتمال وقوع رعد و برق در شهر رشت استفاده شود؛ همچنین مشخص شد با توجه به نتایج مدل‌ها، پارامترهای مؤثر در وقوع رعد و برق به ترتیب اهمیت، رعد و برق در روز گذشته، دما، فشار هوا، رطوبت نسبی و ابرناکی هستند و سایر پارامترها اهمیت کمتری دارند. تعیین مدل ماشین بردار پشتیبان به‌مثابه مدل بهینه پس از ارزیابی با بالاترین عملکرد، حداکثر صحت و کمترین خطا در پیش‌بینی، موجب صرفه‌جویی در زمان برای انجام پژوهش‌های دیگر و انجام اقدام به‌موقع به‌منظور کاهش خسارت در آینده خواهد شد. استفاده از تکنیک داده‌کاوی و پیش‌بینی احتمال وقوع رعد و برق در آینده با استفاده از مدل ماشین بردار پشتیبان به‌منزله مدلی با دقت و صحت حداکثر امکان‌های هواشناسی دقیق‌تر و اقدامات مؤثرتر به‌منظور کاهش خسارت در آینده را فراهم می‌سازد.

### منابع

- جلالی، اوج؛ رسولی، علی‌اکبر؛ ساری صراف، بهروز (۱۳۸۵). توفان‌های تندری و بارش‌های ناشی از آن در محدوده شهر اهر. *جغرافیا و برنامه‌ریزی*، ۱۶ (۲۴)، ۱۸-۳۳.
- خالصی، فریده (۱۳۹۳). واکاوی زمانی توفان‌های تندری در ایران. *آب‌وهواشناسی کاربردی*، ۱ (۱)، ۴۷-۶۰.
- خورشیددوست، علی‌محمد؛ رسولی، علی‌اکبر؛ فخاری واحد، مجتبی (۱۳۹۶). توزیع زمانی و مکانی پدیده رعدوبرق در ایران با استفاده از داده‌های سنجنده ثبت رعد و برق (LIS). *جغرافیا و مخاطرات محیطی*، ۶ (۲۱)، ۸۹-۱۰۷.
- خوشحال دستجردی، جواد؛ قویدل رحیمی، یوسف (۱۳۸۶). شناسایی ویژگی‌های سوانح محیطی منطقه شمال غرب ایران (نمونه مطالعاتی: خطر توفان‌های تندری در تبریز). *فصلنامه مدرس علوم انسانی*، ۱۱ (۵۳)، ۱۰۱-۱۱۵.
- رسولی، علی‌اکبر؛ بوداق جمالی، جواد؛ جلالی، اوج (۱۳۸۶). توزیع زمانی بارش‌های رعد و برقی منطقه شمال غرب ایران. *مجله پژوهشی علوم انسانی دانشگاه اصفهان*، ۱۴ (۲۲)، ۱۵۵-۱۷۰.
- رسولی، علی‌اکبر؛ جوان، خدیجه (۱۳۹۱). تحلیل روند وقوع طوفان‌های رعد و برق در نیمه غربی ایران با کاربرد آزمون‌های ناپارامتری. *مجله فضای جغرافیایی*، ۱۲ (۳۸)، ۱۱۱-۱۲۶.
- قویدل رحیمی، یوسف؛ باغبانان، پرستو؛ فرج‌زاده اصل، منوچهر (۱۳۹۳). تحلیل فضایی مخاطره توفان‌های تندری بهاره ایران. *تحلیل فضایی مخاطرات محیطی*، ۱ (۳)، ۵۹-۷۰.
- مفتاحی نمین، شیدا؛ صلاحی، برومند (۱۳۹۲). تحلیل ویژگی‌های آماری و همدیدی توفان‌های تندری شهرستان آستارا. در:

عبدالله حسن‌زاده، دومین همایش ملی تغییر اقلیم و تأثیر آن بر کشاورزی و محیط‌زیست، (صص. ۲۹۲۰-۲۹۲۹). ارومیه: مرکز تحقیقات کشاورزی و منابع طبیعی استان آذربایجان غربی.

موسوی، محبوبه؛ زرین، آذر؛ مفیدی، عباس؛ حسینی، فاطمه (۱۳۹۶). بررسی ارتباط بین فراوانی وقوع توفان‌های تندری و روند دما در شهر مشهد. *تحقیقات جغرافیایی*، ۳۲(۳)، ۷۵-۸۷.

نصیری قلعه‌بین، سحر؛ صلاحی، برومند؛ رسولی، علی‌اکبر؛ خوش‌اخلاق، فرامرز (۱۳۹۸). تحلیل زمانی مکانی طوفان‌های تندری در دشت اردبیل. *پژوهش‌های جغرافیایی طبیعی*، ۵۱ (۱)، ۱۴۹-۱۶۲.

## References

- Bala, K., Choubey, D. K. & Paul, S. (2017). Soft Computing and Data Mining Techniques for Thunderstorms and Lightning Prediction: A Survey. *International Conference on Electronics Communication and Aerospace Technology ICECA*, 42-46.
- Barnes, S. L. & Newton, C. W. (1982). *Thunderstorms in the synoptic setting, in thunderstorms, In: Kessler, E. (Editor), A Social Scientific and Technological Documentary, Thunderstorm Morphology and Dynamics*. Washington: U.S Deptment of Commerce, D.C 2, 109-171.
- Basak, P., Sarkar, D. & Mukhopadhyay, A. K. (2012). Estimation of Thunderstorm Days from the Radio-sonde Observations at Kolkata (22.53 N, 88.33 E), India during Pre-Monsoon Season: an ANN Based Approach. *Earth Science India*, 5(4), 139-151.
- Blouin, D. K., Flannigan, D. M., Wang, X. & Kochtobajda, B. (2016). Ensemble lightning prediction models for the province of Alberta, Canada. *International Journal of Wildland Fire*, 25 (4), 421-432.
- Burrows, W. R., Price, C. & Wilson, L. J. (2005). Warm season lightning probability prediction for Canada and the northern United States. *Weather Forecasting*, 20, 971-988.
- Chaudhuri, S. & Middey, A. (2013). Nowcasting lightning flash rate and peak wind gusts associated with severe thunderstorms using remotely sensed TRMM-LIS data. *Journal of Remote Sensing*, 34 (5), 1576-1590.
- Chauhan, D. & Thakur, J. (2014). Data Mining Techniques for Weather Prediction: A Review. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2 (8), 2184-2189.
- Court, A. & Griffiths, J. F. (1982). *Thunderstorm climatology, In: Kessler, E. (Editor), Thunderstorms: A Social, Scientific and Technological Documentary, Thunderstorm Morphology and Dynamics*. Washington: U.S Deptment of Commerce, D.C 2, 11-52.
- Ghavidel Rahini, Y., Baghebanan, P. & Farajzadeh, M. (2014). The spatial analysis of hazard of spring thunderstorms in Iran. *Spatial Analysis Environmental Hazards*, 1 (3), 59-70. (In Persian)
- Han, J., Kamber, M. & Pei, J. (2006). *Data Mining Concepts and Techniques*. Second Edition, Burlington, USA: Morgan Kaufmann.
- Hodanish, S. & Wolyn, P. (2012). April, Lightning climatology for the state of Colorado. *Paper presented at the 23 rd International Lightning Detection Conference & 4th International Lightning Meteorology Conference, Broomfield, Colorado, USA*.
- Hou, S., Hou, R., Shi, X., Wang, J. & Yuan, C. (2014). Research on C5.0 Algorithm Improvement and the Test in Lightning Disaster Statistics. *International Journal of Control and Automation*, 7 (1), 181-190.
- Jalali, A., Rasouly, A. A. & Sari Sarraf, B. (2006). Thunderstorms and its rains in Ahar city. *Geography and Planning*, 16 (24), 18-33. (In Persian)
- Jian, C., Gao, J. & Ao, Y. (2016). A new sampling method for classifying imbalanced data based on support vector machine ensemble. *Neurocomputing*, (193), 115-122.
- Khalesi, F. (2014). A temporal analysis of thunderstorms in Iran. *Applied Climatology*, 1 (1), 47-60. (In Persian)
- Khorshiddoust, A. M., Rasouly, A. A. & Vahed, M. F. (2017). Spatio-temporal distribution of lightning phenomenon in Iran using TRMM Lightning Image Sensor (LIS) data. *Geography and Environmental Hazards*, 6 (21), 89-107. (In Persian)

- Khoshhal Dastjerdi, J. & Ghavidel Rahini, Y. (2007). Identify Characteristics of environmental disasters in Northwest Iran (Case Study: Thunderstorm Risk in Tabriz). *Humanities Teacher Quarterly*, 11 (53), 101-115. (In Persian)
- Kodama, Y. M., Okabe, H., Tomisaka, Y., Kotono, K., Kondo, Y. & Kasuya, H. (2007). Lightning frequency and microphysical properties of precipitating clouds over the western North Pacific during winter as derived from TRMM multisensor observations. *Monthly Weather Review*, 135 (6), 2226-2241.
- Kohavi, P. (1998). Glossary of terms, editorial for the special issue on applications of machine learning and the knowledge discovery process. *Machine Learning*, 30 (2-3), 271-274.
- Lambert, W. C., Wheeler, M. & Roeder, W. (2005). Objective lightning forecasting at Kennedy Space Center and Cape Canaveral Air Force Station using cloud-to-ground lightning surveillance system data; Preprints. *Conference on Meteorological Applications of Lightning Data*, (pp. 1-10). San Diego, CA: American Meteorological Society.
- McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T. & Williams, J. K. (2017). Using Artificial Intelligence to Improve Real-Time Decision N-Making for High-Impact Weather. *American Meteorological Society*, 98 (10), 2073-2090.
- Meftahi Namin, S. & Salahi, B. (2013). Analysis of statistical and synoptic characteristics of thunderstorms in Astara city. In: Abdullah Hassanzadeh. *Second International Conference on Climate Change and Impact on Agriculture and Environment*, (pp. 2920-2929). Urumia: Agriculture Research Center and Natural Resources of West Azarbaijan. (In Persian)
- Moosavi, M., Zarrin, A., Mofidi, A. & Hosseini, F. (2017). Investigating the relationship between the frequency of thunderstorms and temperature trend in Mashhad. *Geographical Researches Quarterly*, 32 (3), 75-87. (In Persian)
- Mostajabi, A. H., Finney, D., Rubinstein, M. & Rachidi, F. (2019). Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *Npj Climate and Atmospheric Science*, 2 (1), 1-15.
- Nasiri Ghalebini, S., Salahi, B., Rasouly, A. A. & Khoshakhlagh, F. (2019). Modeling spatial distribution of thunderstorm rainfalls in mountainous area of the Northwest Iran. *Physical Geography Research Quarterly*, 51 (1), 149-162. (In Persian)
- Qiu, T., Zhang, S., Zhou, H., Bai, X. & Liu, P. (2013). Application study of machine learning in lightning forecasting. *Journal of Technology*, 12 (21), 6031-6037.
- Rajeevan, M., Madhulatha, A., Rajasekhar, M., Bhate, J., Kesarkar, A., Kesarkar, A. & Appa Rao, B. V. (2012). Development of a perfect prognosis probabilistic model for prediction of lightning over south-east India. *Journal Earth System Science*, 121 (2), 355-371.
- Rasouly, A. A., Budagh Jamali, J. & Jalali, A. (2007). Spatial distribution of lightning rainfalls in North west Iran. *Humanities of Isfahan University*, 14 (22), 155-170. (In Persian)
- Rasouly, A. A. & Javan, K. H. (2012). Analysis of thunderstorm occurrence trends in the Western part of Iran applying Non-Parametric Statistical tests. *Geographical Space*, 12 (38), 111-126. (In Persian)
- Reap, R. M. (1994). Analysis and prediction of lightning strike distributions associated with synoptic map types over Florida. *Monthly Weather Review*, (122), 1698-1715.
- Zhang, Z., Krawczyk, B., Garcia, S., Rosales-Pérez, A. & Herrera, F. (2016). Empowering one-vs-one decomposition with ensemble learning for multiclass imbalanced data. *Knowledge Based System*, 106, 251-263.
- Zhu, Y. M., Lu, X. X. & Zhou, Y. (2007). Suspended sediment flux modeling with artificial neural network: An example of the longchuanjiang river in the Upper Yangtze Catchment, China. *Geomorphology*, 84 (1), 111-125.



