

امکان سنجی پیش‌بینی باد جستی با استفاده از داده‌های رگرسیون بر اساس مجموع داده‌های میدانی محدود

فریده حبیبی^{*۱}

استادیار، موسسه ژئوفیزیک دانشگاه تهران، تهران، ایران

(دریافت: ۱۳۹۹/۰۸/۲۳، پذیرش: ۱۳۹۹/۱۱/۱۲)

چکیده

این تحقیق با استفاده از اطلاعات ایستگاه خودکار باند فرودگاه مهرآباد در بازه زمانی ژانویه تا ژوئن ۲۰۱۳ و گزارش متار سال ۲۰۱۳ ایستگاه مهرآباد، امکان پیش‌بینی جهت و سرعت باد جستی را با روش رگرسیون بررسی کرده است. داده‌های ایستگاه خودکار از سه سنجنده مستقر در باند به طول ۴۰۰۰ متر و عرض ۴۵ متر گرفته شده است که در راستای جنوب شرق به شمال غرب واقعند. به‌جز جهت و سرعت باد جستی که با حسگرها اندازه‌گیری می‌شوند، تمامی داده‌هایی که سنجنده‌ها اندازه‌گیری کرده‌اند، به بازه ۰/۱-۰/۹ استانداردسازی شدند. ۷۰ درصد کل داده‌ها به‌صورت تصادفی برای آموزش، ۱۵ درصد جهت آزمون و ۱۵ درصد نیز جهت اعتبارسنجی در نظر گرفته شدند که در فایل‌های جداگانه ذخیره و از آنها به جای داده‌های اصلی در طول محاسبات استفاده شده است. تمام کمیت‌های استانداردسازی شده با استفاده از سه روش انتخاب ویژگی شامل روش پی‌درپی پیشرو، روش پسرو و روش اطلاعات متقابل به روش کمترین افزونگی و بیشترین ارتباط پردازش شدند. سپس ویژگی‌های انتخابی، جداگانه، در رگرسیون خطی برای پیش‌بینی جهت و سرعت باد جستی در فصل‌های زمستان و بهار به کار رفتند. نتایج این تحقیق نشان می‌دهد ویژگی‌های انتخابی با روش پسرو برای پیش‌بینی سرعت باد در فصل زمستان مشابه فصل بهار است اما برای جهت باد، کمی متفاوت هستند. ویژگی‌های انتخابی از روش پیشرو برای پیش‌بینی جهت و سرعت باد جستی در فصل زمستان، زیرمجموعه‌ای از ویژگی‌های انتخابی برای فصل بهار هستند. برای این دو فصل، ویژگی‌های انتخابی با اطلاعات متقابل، مشابه است ولی وزن‌های متفاوتی دارد. عملکرد این روش در پیش‌بینی سرعت باد جستی بهتر از پیش‌بینی جهت باد است. روش پیشرو در انتخاب ویژگی برای جهت و سرعت باد جستی، در باند میانی بهینه است. در باند ۱۱ روش پسرو برای پیش‌بینی سرعت و روش پیشرو برای پیش‌بینی جهت باد جستی بهینه است. در باند ۲۹، روش پسرو برای انتخاب ویژگی‌های مرتبط با سرعت و جهت باد جستی بسیار مناسب است. در نهایت، با بررسی خروجی مدل‌ها برای هریک از باندها معادله‌ای برای پیش‌بینی جهت و سرعت باد جستی ارائه شده است.

واژه‌های کلیدی: انتخاب ویژگی پیشرو، پسرو، اطلاعات متقابل، باد جستی، پیش‌بینی، رگرسیون، فرودگاه مهرآباد

امقدمه

پیش‌بینی کمیت‌های هواشناسی هدفی است که از دیرباز پژوهشگران هواشناسی به دنبال آن هستند و برای دستیابی به روش‌های صحیح و کارآمد در زمینه پیش‌بینی، همواره در تلاشند. نقطه شروع اندازه‌گیری کمیت‌های هواشناسی به شکل کنونی، به اختراع فشارسنج در اواسط قرن هفدهم برمی‌گردد. ایستگاه‌های هواشناسی همدیدی در سال ۱۸۱۶ تأسیس شدند، اما اقدام برای صدور پیش‌بینی هواشناسی تا اواسط قرن نوزدهم به تأخیر افتاده است (بلواشتاین، ۱۹۹۲). در این ایستگاه‌ها در ابتدا فقط فشار، دما و میزان ابرناکی بررسی و پیش‌بینی کوتاه‌مدت صادر می‌شد، ولی از اواسط قرن بیستم با تجهیز ایستگاه‌های هواشناسی، تعداد کمیت‌هایی که اندازه‌گیری می‌شوند، افزایش یافت. به این ترتیب، پژوهشگران هواشناسی با حجم عظیمی از اطلاعات و داده مواجه شدند که هوش و قابلیت‌های انسانی، توان و زمان بررسی کامل آنها را نداشت. این حجم زیاد اطلاعات ذخیره‌شده، ایشان را بر آن داشت که به دنبال کشف دانش از داده یا همان داده‌کاوی (data mining) باشند تا بتوانند مدلی ارائه دهند که مفهومی فراتر از این داده‌های خام داشته باشد (رضائی یوسفی، ۱۳۸۶). واپازی یا رگرسیون، که در لغت به معنی بازگشت است، یکی از روش‌های رایج در داده‌کاوی است (مقصود و اچوآن، ۲۰۱۶) و مفهوم کاربردی آن، پی بردن به رفتار یک متغیر با استفاده از رفتار متغیر یا متغیرهای دیگر است. زمانی که داده‌های متغیرهای یک پدیده به صورت یک سری زمانی در دسترس باشد، با استفاده از رگرسیون نه تنها می‌توان قدرت نفوذ پیش‌بینی‌کننده‌های فردی را بر متغیر وابسته منفرد نشان داد، بلکه می‌توان چگونگی تغییر متغیر وابسته را با تغییر پیش‌بینی‌کننده‌ها نشان داد. همچنین می‌توان یک مدل خوب برای پیش‌بینی روندها و مقادیر کمیت در چند گام زمانی آینده ارائه کرد (هزرا و گوگتای، ۲۰۱۷). پس

رگرسیون در هر زمینه‌ای که با داده‌های سری زمانی سروکار داشته باشد، کاربرد دارد و به همین دلیل نیز امروزه، تحلیل رگرسیونی، یک فن آماری با بیشترین و وسیع‌ترین کاربرد بین روش‌های آماری است.

بررسی و مدل‌سازی صحیح پدیده‌های جوی مستلزم شناخت کامل کمیت‌های تأثیرگذار است. گرچه می‌توان از دانش قبلی فرد با تجربه کمک گرفت، این دانش در افراد با تجربه متفاوت است. برای گریز از این مشکل، معمولاً از روش‌های آماری و هوش مصنوعی کمک می‌گیرند. استفاده از این روش‌ها نیز مستلزم انتخاب متغیرهای ورودی مناسب برای مدل‌سازی آن پدیده است (رضائی یوسفی، ۱۳۸۶). امروزه، این چالش را چالش انتخاب و استخراج ویژگی می‌نامند. هدف از انتخاب ویژگی، یافتن بهترین زیرمجموعه از مجموعه ویژگی‌ها است که کمترین خطای کلیت‌بخشی را به همراه داشته باشد (ورگارا و استیوز، ۲۰۱۴).

در آمار، رگرسیون دو بخش پارامتری و ناپارامتری دارد. در مدل‌هایی که بر مبنای پارامتری استوارند، محاسبات به لحاظ شناخته بودن نوع ارتباط بین متغیرهای وابسته و مستقل، سریع‌تر انجام می‌شود، اما به ایجاد فرضیات قوی در خصوص ماهیت توزیع داده‌ها نیاز است. به عبارتی، این مدل‌ها داده‌محور هستند (مورفی، ۲۰۱۲ و فتاحی، ۲۰۱۱). در مقابل، مدل‌های ناپارامتری، کمترین فرضیات را در تحلیل توزیع داده دارند و در نتیجه، انعطاف‌پذیرتر هستند، اما اگر مجموعه داده‌ها بزرگ باشد، اغلب نمی‌توان محاسبات را کنترل کرد (مورفی، ۲۰۱۲).

هدف از رگرسیون، یافتن رابطه‌ای ریاضی برای تعیین کمیت متغیر وابسته (Y) با استفاده از متغیر یا متغیرهای مستقل (X) است. رگرسیون به صورت‌های مختلف اجرا می‌شود. رگرسیون رایج کلاسیک و آماری دربرگیرنده مدل‌های خطی است که می‌تواند ساده $(f(X)=aX+b)$ یا

جدول ۱. انواع توابع در رگرسیون غیرخطی. در این روابط نماد * ضرب و ** به مفهوم توان است (حبیبی، آ، ۱۳۹۵ ص. ۱۲۲).

Name	Model expression
Asymptotic Regression/Growth Model	$b1 + b2 * \exp(b3 * x)$
Logistic Population Growth Model	$b1 / (1 + \exp(b2 + b3 * x))$
Asymptotic Regression/Decay Model	$b1 - (b2 * (b3 ** x))$
Saturation Model	$b1 * \exp(b2 * x)$
Density	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
Log-Modified	$(b1 + b3 * x) ** b2$
Log-Logistic	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Metcherlich Law of Diminishing Returns	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
Ratio of Cubics	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
Ratio of Quadratics	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x ** b4)$
Yield Density	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

جمع‌آوری داده‌ها (Global Data Assimilation and Prediction System, GDAPS) به مجموعه شبکه تودرتو وارد و مدل برای مدت زمان سی ساعت اجرا شده است. البته شبکه تودرتو در این تحقیق شامل پنج شبکه به ترتیب با ابعاد ۴/۴ کیلومتر و ۱/۵ کیلومتر (هر دو با نقاط شبکه ۲۰۰×۲۰۰)، ۷۵۰ متر و ۵۰۰ متر (هر دو با نقاط شبکه ۳۰۰×۳۰۰) و ۳۰۰ متر (با نقاط شبکه ۴۰۰×۴۰۰) با تفکیک بیشتر از ۳۰۰ متر است. جو نیز ۷۰ ترازی فرض شده است. در این مطالعه دو نمونه از باد جستی قوی برای پیش‌بینی انتخاب شده بود که یکی در ۳۰ اوت ۲۰۱۶ و دیگری در ۲ اکتبر ۲۰۱۶ در این منطقه اتفاق افتاده بود. نتایج این تحقیق نشان می‌دهد علی‌رغم بهبود در مدل-IIA 300 متر، پیش‌بینی باد جستی بهبود چندانی نیافته است؛ زیرا پوشش گیاهی زمین، سطح اقیانوس مجاور و شرایط آغازین مدل در قیاس با شرایط مرزی، نقش عمده‌ای در ضعف نتایج این پیش‌بینی داشتند. به گفته ایشان، در مطالعات بعدی، با وارد کردن داده‌های اضافی، نتایج این تحقیق بهبود خواهد یافت. حبیبی (۱۳۹۸) با استفاده از دو روش انتخاب ویژگی اطلاعات متقابل و جستجوی

چندگانه، غیرخطی، چندجمله‌ای یا ترکیبی از حالت‌های عددی و توابع سینوسی به صورت جدول ۱ باشد (حبیبی، آ، ۱۳۹۵). در ادامه، برخی از مطالعات انجام شده در زمینه هواشناسی با استفاده از روش‌های انتخاب ویژگی و رگرسیون معرفی می‌شوند. پراسانا و همکاران (۲۰۲۰) با استفاده از مدل UM نسخه ۱۰/۴ و برنامه جانبی مرکزی نسخه ۹/۰ به پیش‌بینی بادجستی پرداختند. آنها از داده‌های فرودگاه بین‌المللی اینچئون(S113) استفاده کردند که پنج ایستگاه هواشناسی خودکار دارد. این ایستگاه‌ها محدوده‌ای به وسعت ۱۲۰ کیلومترمربع را پوشش می‌دهند. از داده‌های ایستگاه شماره ۱۱۳ که تقریباً در وسط این ناحیه قرار دارد، فقط برای اعتبارسنجی استفاده شده است. در اتصال بین مدل‌ها، متغیرهای پیش‌یابی جو (بردارهای باد، دمای پتانسیل، چگالی، فشار و نسبت آمیختگی جرم، بخار آب، مایع و میعانات یخ‌زده ابر، باران و هواویزها) و تمام متغیرهای پیش‌یابی سطحی و جنب‌سطحی مانند دما و رطوبت خاک نیز ساعت به ساعت از سامانه جهانی پیش‌بینی و

می‌گذارد، اما سرعت باد، که یک عامل هواشناختی است، بر بار الکتریکی تأثیر چندانی ندارد.

موستریس و همکاران (۲۰۱۲) بیشینه غلظت سطح ازن برای ۲۴ ساعت آینده در منطقه آتن را با استفاده از مدل‌های رگرسیون خطی چندگانه و بر اساس رویکرد شبکه عصبی مصنوعی پیش‌بینی کردند. آگبو و همکاران (۲۰۱۳) با تحلیل رگرسیونی مرتبه اول و دوم و همبستگی چندگانه توانستند تابش خورشیدی در ساعات صبح را برای منطقه لاپای در نیجریه تخمین بزنند و استفاده از این روش را برای مناطقی با اقلیم مشابه اقلیم لاپای توصیه کردند.

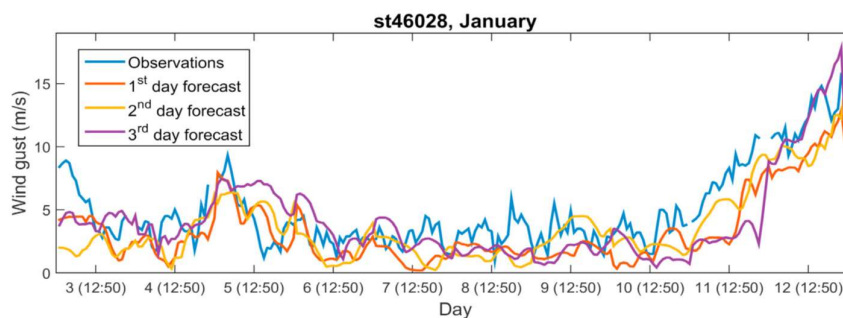
روندی و فرانک (۲۰۰۴) برای یافتن روابط بین مدهای موج برهم‌کنشی که با فرکانس‌های مختلف مشخص می‌شوند، از مدل‌های رگرسیون خطی چندگانه با جملات توانی غیرخطی جهت تحلیل گردش‌های جوئی استفاده و برای پیش‌بینی بی‌هنجاری‌های آب بارش شو بین‌فصلی که حرکت غرب‌سو دارند، مدلی ارائه کردند.

رضایی و همکاران (۱۳۸۷) با استفاده از پارامترهای هواشناسی به‌عنوان متغیرهای مستقل در تحلیل رگرسیون خطی و غیرخطی، برای تعیین جریان نشتی آلوده روی مقره‌های شبکه توزیع برق در مناطق آلوده سواحل جنوبی، مدل رگرسیون خطی و غیرخطی ارائه کردند. پاتلاکاس و همکاران (۲۰۱۷) با استفاده از مدل‌های پیش‌بینی عددی وضع هوا و پس‌پردازش آماری توانستند روشی جهت برآورد سرعت باد جستی سطحی در ۲۴، ۴۸ و ۷۲ ساعت آینده برای استفاده در مزارع بادی نزدیک به ساحل ارائه دهند. در این روش برای تخمین‌های آماری از فیلترکالمن استفاده شده است. آنها با ارائه شکل ۱ نشان دادند که روش یادشده برای پیش‌بینی ۲۴ ساعته سرعت باد جستی موفق‌تر است و استفاده از فیلترکالمن نیز در رفع خطاهای سیستماتیک مدل عددی مفید بوده است. این مدل سرعت باد را برای مکان‌های نزدیک به خط ساحلی، کم برآورد

پی‌درپی پیشرو شناور با الگوریتم طبقه‌بندی k نزدیک‌ترین همسایگان، ابتدا ویژگی‌های مناسب برای پیش‌بینی جهت و سرعت بادجستی در ایستگاه فرودگاهی را یافته و سپس با استفاده از شبکه عصبی پرسپترون چندلایه به پیش‌بینی آنها پرداخته است. در این بررسی ویژگی‌های انتخابی برای جهت باد از تنوع بیشتری برخوردار است. بعد از اعمال ویژگی‌های انتخابی به شبکه عصبی پرسپترون چندلایه در حالت‌های مختلف شبکه، نتایج خروجی مدل برای پیش‌بینی جهت و سرعت باد جستی مقایسه و بهترین مدل انتخاب شده است. بر اساس این مطالعه، عملکرد شبکه عصبی پرسپترون چندلایه در پیش‌بینی سرعت باد جستی بهتر از پیش‌بینی جهت باد جستی است. در تحقیق دیگری، حبیبی (۱۳۹۵) با استفاده از گزارش‌های همدیدی و متار ایستگاه بندرعباس و روش رگرسیون لجستیک، رابطه‌ای را برای پیش‌بینی بارش ارائه کرده و نشان داده است که دمای نقطه شبنم و دید افقی، بر میزان احتمال بارش، اثر کاهندگی و بر میزان ابرناکی آسمان، اثر افزونگی دارند.

عرب عامری و همکاران (۱۳۹۳) از سری زمانی داده‌های متار ایستگاه مهرآباد و اعمال روش‌های مختلف انتخاب ویژگی برای پیش‌بینی سرعت باد استفاده کردند. آنها معتقدند که روش انتخاب ویژگی پسرو، در بین روش‌هایی که به‌کاربردند بهترین است. همچنین اگر ورودی به شبکه عصبی پرسپترون چندلایه شامل دما، دمای نقطه شبنم، فشار و سوگیری (bias) باشد، شبکه می‌تواند سرعت باد را برای یک ساعت آینده به‌خوبی پیش‌بینی کند.

گلادیس و کوچتا (۲۰۰۸) از درخت‌های رگرسیون برای تحلیل بار الکتریکی و شناسایی عوامل مؤثر بر تقاضای برق و تعیین ماهیت نفوذ عنصر مهم از نظر مدیریت انرژی مؤثر در لهستان استفاده کردند و نتیجه گرفتند که شرایط جوئی به‌طور چشمگیری بر میزان بار الکتریکی تأثیر



شکل ۱. سری زمانی داده‌های بویه (خط آبی) و پیش‌بینی‌های مدل برای ۲۴ ساعت (خط قرمز)، ۴۸ ساعت (خط زرد) و ۷۲ ساعت (خط بنفش) (پاتلاکاس و همکاران، ۲۰۱۷).

سنجنده میانی آنها را اندازه‌گیری می‌کند، بقیه شامل کمیت‌های اندازه‌گیری شده از سه سنجنده مختلف هستند. متغیرهای ردیف ۵۹ تا ۶۴ مربوط به جهت و سرعت باد جستی هستند که سنجنده‌ها اندازه‌گیری می‌کنند و برای مدل‌های اجرا شده در این تحقیق حکم هدف را دارند.

داده‌های متار: این داده‌ها شامل جهت و سرعت باد (میانگین ده دقیقه‌ای باد در ارتفاع ۱۰ متری از سطح زمین)، جهت و سرعت باد جستی، درجه حرارت و دمای نقطه شبنم، فشار ایستگاه، هوای حاضر، ابرناکی و به‌ویژه میزان و ارتفاع تشکیل ابر کومه‌ای هستند که برای چهار فصل سال (بر اساس تاریخ شمسی) پردازش شده‌اند. نتایج این بررسی در جدول ۳ و نمودارهای شکل ۲ ارائه شده است. از آنجا که مرتبه گزارش‌های مربوط به ارتفاع تشکیل ابرها نسبت به بقیه پدیده‌ها بزرگ‌تر است، برای وضوح بیشتر، نمودار در دو بخش ترسیم شده است. مقایسه فراوانی وقوع پدیده‌ها در فصل‌های مختلف نشان می‌دهد در این بازه زمانی، بیشترین ناپایداری‌ها (شکل‌گیری ابر کومه‌ای باران و وقوع طوفان‌های تندی) در فصل‌های بهار و زمستان رخ داده است.

داده‌های سنجنده‌های خودکار: داده‌های باند فرودگاه مهرآباد که هم‌زمان و در فواصل زمانی یک دقیقه‌ای گردآوری شده بودند، ابتدا پیش‌پردازش شدند؛ یعنی با حذف گزارش‌های نادرست (داده‌های پرت) و تصحیح

می‌کند؛ چون در مرزهای زمین-آب، بازنمایی ضعیفی دارد.

هدف این تحقیق، یافتن ارتباط بین کمیت‌های اندازه‌گیری شده در ایستگاه خودکار فرودگاهی مهرآباد با جهت و سرعت باد جستی با استفاده از روش‌های داده‌کاوی و ارائه مدلی برای پیش‌بینی آنها با روش رگرسیون است.

۲ داده‌ها و روش تحقیق

داده‌های استفاده شده در این تحقیق شامل داده‌های گزارش متار ایستگاه مهرآباد متعلق به سال ۲۰۱۳ میلادی است که در فواصل زمانی سی دقیقه‌ای گزارش و از وبگاه <https://www.ogimet.com/index.phtml.en> تهیه شده‌اند. علاوه بر این، داده‌های سنجنده‌های خودکار باند فرودگاه مهرآباد در همان سال نیز از سازمان هواشناسی کشور تهیه شده است. این باند به طول ۴۰۰۰ متر و عرض ۴۵ متر در راستای جنوب شرق به شمال غرب قرار دارد. سنجنده شماره ۲۹ در منتهی‌الیه شمال غربی باند، سنجنده شماره ۱۱ در منتهی‌الیه جنوب شرقی باند و سنجنده میانی به فاصله ۶۰۰ متری وسط باند در جهت شمال آن قرار دارد. این سنجنده‌ها کمیت‌های معرفی شده در جدول ۲ را هم‌زمان و در فواصل زمانی یک دقیقه‌ای اندازه‌گیری می‌کنند. البته به‌جز متغیرهای ردیف ۵۲ تا ۵۸ که فقط

جدول ۲. کمیت‌های اندازه‌گیری شده در ایستگاه خودکار فرودگاه مهرآباد.

کد	تفسیر	کد	تفسیر
۱	WSINS29	CW2A_Mid	سرعت باد لحظه‌ای در باند ۲۹
۲	WS2M29	HW2A_Mid	حداقل سرعت باد در دو دقیقه در باند ۲۹
۳	WS2A29	WSINS11	متوسط سرعت باد در دو دقیقه در باند ۲۹
۴	WS2X29	WS2M11	حداکثر سرعت باد در دو دقیقه در باند ۲۹
۵	WS10M29	WS2A11	حداقل سرعت باد در ده دقیقه در باند ۲۹
۶	WS10A29	WS2X11	متوسط سرعت باد در ده دقیقه در باند ۲۹
۷	WS10X29	WS10M11	حداکثر سرعت باد در ده دقیقه در باند ۲۹
۸	WDINS29	WS10A11	جهت باد لحظه‌ای در باند ۲۹
۹	WD2M29	WS10X11	حداقل جهت باد در دو دقیقه در باند ۲۹
۱۰	WD2A29	WDINS11	متوسط جهت باد در دو دقیقه در باند ۲۹
۱۱	WD2X29	WD2M11	حداکثر جهت باد در دو دقیقه در باند ۲۹
۱۲	WD10M29	WD2A11	حداقل جهت باد در ده دقیقه در باند ۲۹
۱۳	WD10A29	WD2X11	متوسط جهت باد در ده دقیقه در باند ۲۹
۱۴	WD10X29	WD10M11	حداکثر جهت باد در ده دقیقه در باند ۲۹
۱۵	WDVAR29	WD10A11	انحراف جهت در ده دقیقه گذشته در باند ۲۹
۱۶	CW2A29	WD10X11	مؤلفه‌های جهت باد در راستای باند پرواز ۲۹
۱۷	HW2A29	WDVAR11	مؤلفه‌های سرعت باد در راستای باند پرواز ۲۹
۱۸	WSINS_Mid	CW2A11	سرعت باد لحظه‌ای در باند میانی
۱۹	WS2M_Mid	HW2A11	حداقل سرعت باد در دو دقیقه در باند میانی
۲۰	WS2A_Mid	P0	متوسط سرعت باد در دو دقیقه در باند میانی
۲۱	WS2X_Mid	QNH	حداکثر سرعت باد در دو دقیقه در باند میانی
۲۲	WS10M_Mid	QFF3H	حداقل سرعت باد در ده دقیقه در باند میانی
۲۳	WS10A_Mid	QFFT	متوسط سرعت باد در ده دقیقه در باند میانی
۲۴	WS10X_Mid	TEMP	حداکثر سرعت باد در ده دقیقه در باند میانی
۲۵	WDINS_Mid	RH	جهت باد لحظه‌ای در باند میانی
۲۶	WD2M_Mid	DEWPOINT	حداقل جهت باد در دو دقیقه در باند میانی
۲۷	WD2A_Mid	GUST_SPEED29	متوسط جهت باد در دو دقیقه در باند میانی
۲۸	WD2X_Mid	GUST_DIR29	حداکثر جهت باد در دو دقیقه در باند میانی
۲۹	WD10M_Mid	GUST_SPEED_Mid	حداقل جهت باد در ده دقیقه در باند میانی
۳۰	WD10A_Mid	GUST_DIR_Mid	متوسط جهت باد در ده دقیقه در باند میانی
۳۱	WD10X_Mid	GUST_SPEED11	حداکثر جهت باد در ده دقیقه در باند میانی
۳۲	WDVAR_Mid	GUST_DIR11	انحراف جهت در ده دقیقه گذشته در باند میانی

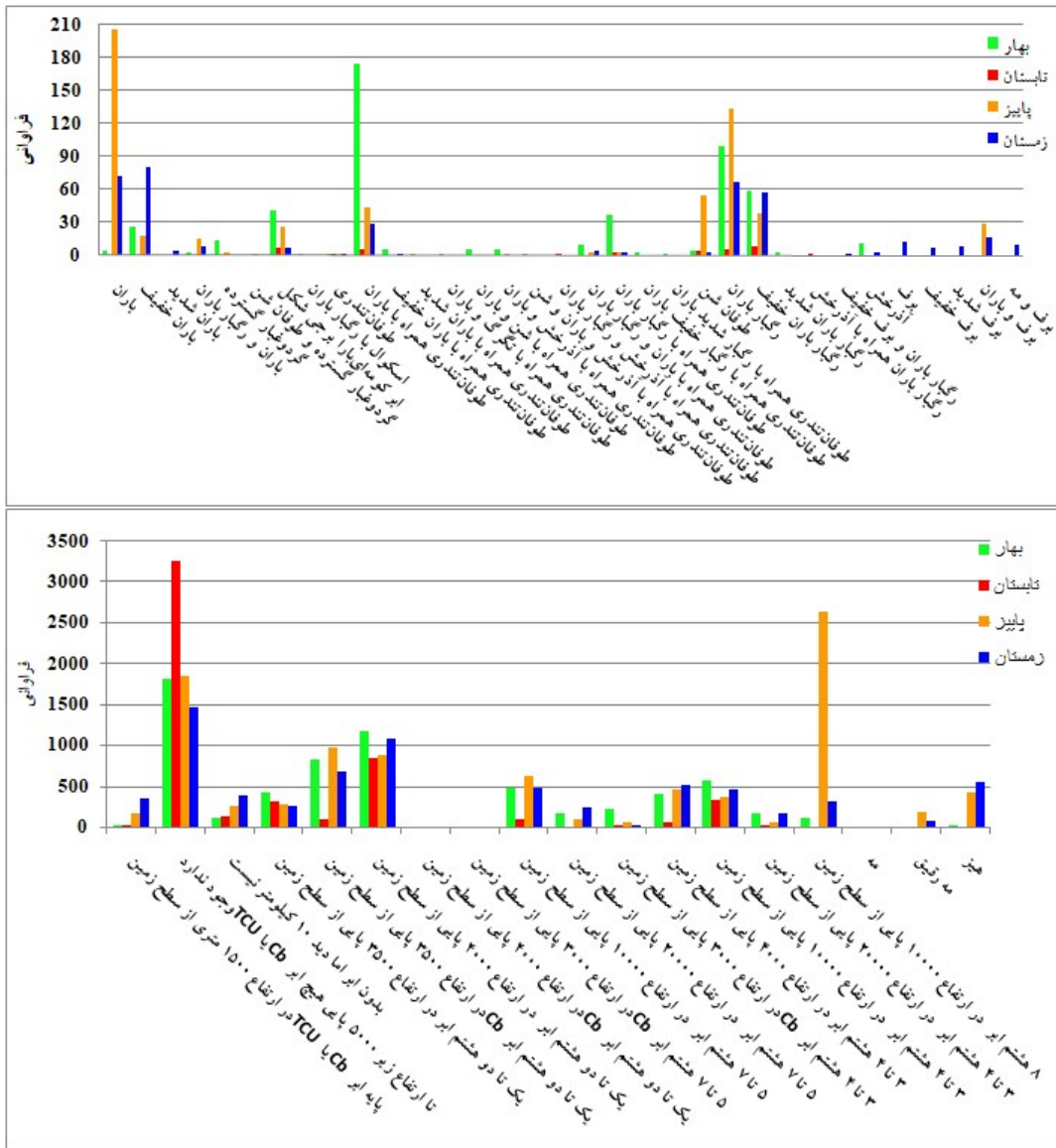
ماه، روز، ساعت و دقیقه است. شش ستون آخر ماتریس نیز در برگیرنده جهت و سرعت باد جستی است که سه سنجنده اندازه‌گیری کرده‌اند. سپس برای ادامه بررسی، داده‌های بازه زمانی ۲۲ دسامبر تا ۲۲ ژوئن ۲۰۱۳ بر اساس نتایج بررسی گزارش‌های متار ایستگاه انتخاب شدند.

داده‌های موردی گزارش‌نشده با استفاده از روش میانگین‌گیری از چهار داده همسایگی، در مجموع ۲۱۹۹۵۷×۶۴ نمونه باقی‌ماند که در یک ماتریس ۲۱۹۹۵۷×۶۴ قرار داده شدند. پنج ستون نخست این ماتریس مربوط به زمان اندازه‌گیری کمیت‌ها است که به ترتیب شامل سال،

۳ روش‌های انتخاب ویژگی

از آنجا که کمیت‌های اندازه‌گیری شده از مرتبه یکسان نیستند، برای افزایش دقت محاسبات و دستیابی به نتایج بهتر، ابتدا از روش هنجارسازی min-max استفاده شد. مزیت این روش در آن است که تمام روابط بین داده‌ها را حفظ می‌کند و به معرفی سوگیری نیاز نیست. هنجارسازی داده‌ها در این روش از رابطه زیر محاسبه شده است (هان و همکاران، ۲۰۱۲):

برای تحلیل هر پدیده‌ای با استفاده از داده‌های جمع‌آوری شده، ابتدا باید بُعد مشاهدات را طوری کاهش داد که با کمترین اطلاعات از دست رفته همراه باشد؛ به عبارتی، بررسی به متغیرهایی اختصاص یابد که از مشاهده‌ای به مشاهده دیگر زیاد تغییر نکنند تا بتوان آنها را به‌عنوان ثابت‌های مسئله در نظر گرفت (حبیبی، ۱۳۹۸).



شکل ۲. فراوانی وقوع پدیده‌ها در فصل‌های مختلف در سال ۲۰۱۳.

جدول ۳. بررسی گزارش‌های متار سال ۲۰۱۳ میلادی.

کد متار	تفسیر هواشناسی	فراوانی پدیده‌ها			
		بهار (۲۱ ژوئن تا ۲۱ مارس)	تابستان (۲۲ ژوئن تا ۲۲ سپتامبر)	پاییز (۲۲ سپتامبر تا ۲۱ دسامبر)	زمستان (۲۲ دسامبر تا ۲۰ مارس)
RA	باران	۴	-	۲۰۶	۷۲
-RA	باران خفیف	۲۵	-	۱۸	۸۰
+RA	باران شدید	-	-	-	۴
RA/SHRA	باران و رگبار باران	۲	-	۱۴	۸
DU	گردوغبار گسترده	۱۳	-	۲	-
SA/DU	گردوغبار گسترده و طوفان شن	-	-	۱	-
TCU	ابر کومه‌ای باران برجی شکل	۴۰	۷	۲۵	۷
SQ_SHRA	اسکوال با رگبار باران	۱	-	-	-
LT	آذرخش	۱۱	-	-	۲
SN	برف	-	-	-	۱۲
-SN	برف خفیف	-	-	-	۷
+SN	برف شدید	-	-	-	۸
SN/RA	برف و باران	-	-	۲۹	۱۶
SN/FG	برف و مه	-	-	-	۹
TS	طوفان تندری	۱	۱	۱	۱
TS/RA	طوفان تندری همراه با باران	۱۷۴	۵	۴۴	۲۸
TS/-RA	طوفان تندری همراه با باران خفیف	۵	-	-	۱
TS/+RA	طوفان تندری همراه با باران شدید	۱	-	-	-
TS/RA/TSGR	طوفان تندری همراه با تگرگ و باران	۱	-	-	-
TS/RA/SA	طوفان تندری همراه با شن و باران	۵	-	-	-
LT/TS/RA	طوفان تندری همراه با آذرخش و باران	۵	-	۱	-
LT/TS/RA/SA	طوفان تندری همراه با آذرخش و باران و شن	۱	-	-	-
LT/TS/SHRA	طوفان تندری همراه با آذرخش و رگبار باران	-	۱	-	-
TS/RA/SHRA	طوفان تندری همراه با باران و رگبار باران	۹	-	۳	۴
TS/SHRA	طوفان تندری همراه با رگبار باران	۳۶	۳	۲	۲
TS/-SHRA	طوفان تندری همراه با رگبار خفیف باران	۲	-	-	-
TS/+SHRA	طوفان تندری همراه با رگبار شدید باران	۱	-	-	-
SA	طوفان شن	۴	۴	۵۴	۲
SHRA	رگبار باران	۱۰۰	۵	۱۳۴	۶۷
-SHRA	رگبار باران خفیف	۵۹	۸	۳۸	۵۷
+SHRA	رگبار باران شدید	۲	-	-	-
LT/SHRA	رگبار باران همراه با آذرخش	-	۱	-	-
SHRA/-SN	رگبار باران و برف خفیف	-	-	-	۱
FG	مه	-	-	-	۸
BR	مه رقیق	-	-	-	۷۳
HZ	هیز	۲۴	۱۱	۴۲۷	۵۵۲
NSC	پایه ابر Cb یا TCU در ارتفاع ۱۵۰۰ متری از سطح زمین	۲۶	۱۵	۱۶۳	۳۵۲
CAVOK	تا ارتفاع زیر ۵۰۰۰ پایی هیچ ابر Cb یا TCU وجود ندارد	۱۸۱۸	۳۲۵۵	۱۸۴۹	۱۴۶۵
SKC	بدون ابر اما دید ۱۰ کیلومتر نیست	۱۱۲	۱۲۷	۲۵۹	۳۸۹
FEW035	یک تا دو هشتم ابر در ارتفاع ۳۵۰۰ پایی از سطح زمین	۴۱۷	۳۰۵	۲۷۴	۲۶۲

ادامه جدول ۳

فراوانی گزارش ابر				تفسیر هواشناسی	کد متار
زمستان	پاییز	تابستان	بهار		
(۲۲ دسامبر تا ۲۰ مارس)	(۲۲ سپتامبر تا ۲۱ دسامبر)	(۲۲ ژوئن تا ۲۲ سپتامبر)	(۲۱ مارس تا ۲۱ ژوئن)	یک تا دو هشتم ابر Cb در ارتفاع ۳۵۰۰ پایی از سطح زمین	FEW035CB
۶۸۶	۹۷۸	۹۸	۸۳۲	یک تا دو هشتم ابر در ارتفاع ۴۰۰۰ پایی از سطح زمین	FEW040
۱۰۹۰	۸۹۰	۸۴۴	۱۱۸۰	یک تا دو هشتم ابر Cb در ارتفاع ۴۰۰۰ پایی از سطح زمین	FEW040CB
--	--	--	۲	۵ تا ۷ هشتم ابر Cb در ارتفاع ۳۰۰۰ پایی از سطح زمین	BKN030CB
--	۲	--	--	۵ تا ۷ هشتم ابر در ارتفاع ۱۰۰۰۰ پایی از سطح زمین	BKN100
۴۷۹	۶۲۰	۱۰۱	۴۷۸	۵ تا ۷ هشتم ابر در ارتفاع ۲۰۰۰۰ پایی از سطح زمین	BKN200
۲۵۱	۸۶	۲	۱۷۱	۳ تا ۴ هشتم ابر Cb در ارتفاع ۳۰۰۰ پایی از سطح زمین	SCT030CB
۲۸	۵۳	۱۷	۲۱۷	۳ تا ۴ هشتم ابر در ارتفاع ۴۰۰۰ پایی از سطح زمین	SCT040
۵۲۶	۴۵۳	۵۱	۴۰۰	۳ تا ۴ هشتم ابر در ارتفاع ۱۰۰۰۰ پایی از سطح زمین	SCT100
۴۶۸	۳۶۹	۳۴۰	۵۶۷	۳ تا ۴ هشتم ابر در ارتفاع ۲۰۰۰۰ پایی از سطح زمین	SCT200
۱۶۷	۵۷	۲۶	۱۷۶	۸ هشتم ابر در ارتفاع ۱۰۰۰۰ پایی از سطح زمین	OVC100
۳۲۵	۲۶۳۳	--	۱۱۸		

۳-۱ انتخاب ویژگی با روش جستجوی پی‌درپی

پیشرو

در روش جستجوی پی‌درپی پیشرو ابتدا ویژگی مورد پیش‌بینی به داخل یک مجموعه تهی مانند S منتقل می‌شود. سپس ویژگی‌های باقیمانده، یک‌به‌یک به مجموعه اضافه و ارزیابی می‌شوند. پس از آن ویژگی‌ای که کمترین خطای ریشه میانگین مربعات (RMSE) را ایجاد کند، به مجموعه S اضافه می‌شود. این عملیات تا جایی ادامه می‌یابد که RMSE مدل نسبت به حالت‌های قبلی افزایش یابد. در این مرحله ویژگی‌های موجود در داخل مجموعه S ویژگی‌هایی هستند که کمترین RMSE را فراهم کرده‌اند و می‌توان از آنها در محاسبات استفاده کرد (وروریدیز و کاتروپلوس، ۲۰۰۵). برای انتخاب ویژگی‌های مناسب برای هر فصل، ابتدا روش برای داده‌های هر فصل اجرا شد. نتایج این روش برای جهت و سرعت باد جستی در هر سه باند در جدول ۴ و به صورت نموداری در شکل ۳ برای باند ۲۹ ارائه شده است

$$x_{i,a \text{ to } b} = a + \frac{(x_i - x_{\text{Min}})(b-a)}{(x_{\text{Max}} - x_{\text{Min}})} \quad (1)$$

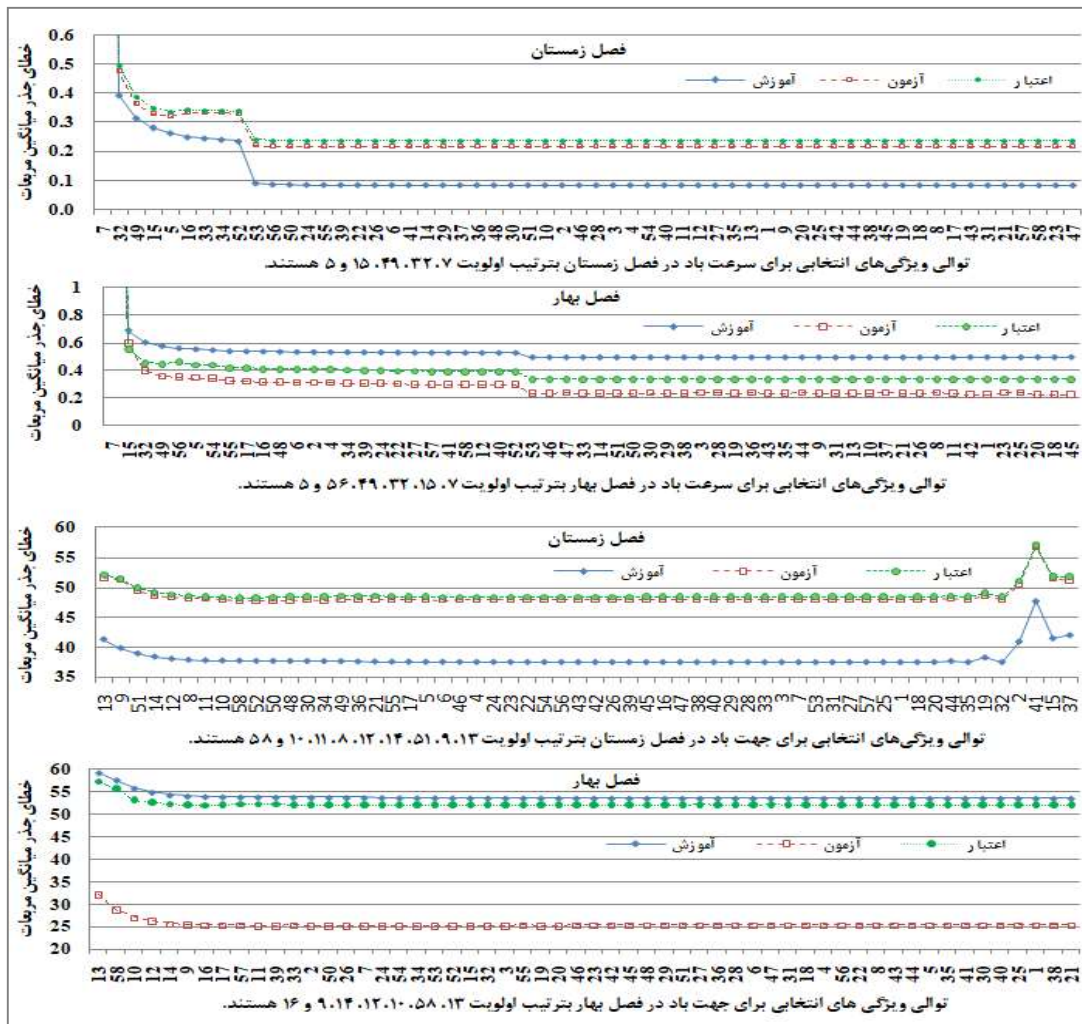
که x_{Min} ، x_{Max} به ترتیب داده‌های ورودی واقعی، کمینه و بیشینه هستند. a و b مقادیر کمینه و بیشینه هدف هستند که در اینجا به ترتیب ۰/۱ و ۰/۹ انتخاب شدند. $x_{i,a \text{ to } b}$ داده i ام هنگام‌سازی شده در فاصله a تا b است؛ بنابراین تمامی داده‌ها با این رابطه به بازه ۰/۹-۰/۱ آورده شدند. در مرحله آخر، با استفاده از روش‌های رایج در انتخاب ویژگی مانند روش جستجوی پی‌درپی پیشرو، پسر و معیار اطلاعات متقابل سعی شد تا ویژگی‌هایی انتخاب شوند که بیشترین تأثیر را در تعیین مقدار کمیت مورد بررسی دارند. در این تحقیق ۷۰ درصد داده‌ها جهت آموزش، ۱۵ درصد جهت آزمون و ۱۵ درصد نیز جهت اعتبارسنجی شبکه در نظر گرفته شدند که به طور تصادفی از بین داده‌ها انتخاب و در فایل‌های جداگانه ذخیره شدند. در طول محاسبات نیز برای اینکه مقایسه نتایج در شرایط یکسان باشد، از این سه فایل به جای داده‌های اصلی استفاده شده است.

جدول ۴. بخشی از خطای جذر میانگین مربعات (RMSE) در انتخاب ویژگی برای جهت و سرعت باد جستی در باندهای فرودگاه مهرآباد. ویژگی‌های انتخابی مناسب برای هر روش با زمینه زرد مشخص شده است.

مقادیر خطای جذر میانگین مربعات (RMSE) در انتخاب ویژگی												روش جستجو	فصل
باند ۲۹				باند میانی				باند ۱۱					
ویژگی	اعتبارسنجی	آزمون	آموزش	ویژگی	اعتبارسنجی	آزمون	آموزش	ویژگی	اعتبارسنجی	آزمون	آموزش		
۵۳	۴۵/۴۳۰۱	۴۶/۱۴۴۳	۴۶/۳۶۰۱	۵۳	۴۸/۳۹۷۰	۴۸/۳۳۱۱	۴۸/۹۲۴۴	۴۸	۴۷/۲۵۴	۴۸/۸۶۷	۴۹/۷۲۸	جهت باد	بهار
۱۴	۴۵/۶۴۸۷	۴۶/۵۴۳۸	۴۶/۷۶۱۴	۳۱	۴۹/۰۷۲۰	۴۸/۹۹۹۳	۴۹/۴۸۴۴	۴۳	۴۷/۴۶۶	۴۹/۰۱۷	۴۹/۹۰۵		
۱۲	۴۶/۱۱۳۳	۴۷/۴۱۵۶	۴۷/۶۴۴۳	۲۹	۴۹/۵۷۲۱	۴۹/۴۷۵۵	۵۰/۰۳۹۱	۴۴	۴۸/۴۶۳	۴۹/۳۵۴	۵۰/۶۰۴		
۱۰	۴۸/۶۵۷۱	۴۸/۹۶۶	۴۹/۱۶۶۴	۲۷	۵۱/۱۷۸۸	۵۱/۱۲۰۹	۵۱/۸۸۳۷	۴۶	۵۰/۸۳۶	۵۱/۹۸۱	۵۳/۱۵۷		
۵۲	۵۱/۳۱۵۴	۵۱/۴۵۴۳	۵۱/۶۴۴	۵۰	۵۴/۰۹۳۸	۵۴/۰۶۱۴	۵۴/۵۲۹۸	۵۱	۵۳/۳۹۷	۵۴/۳۹۶	۵۵/۳۳۶	سرعت باد	بهار
۱۳	۲۲۷/۳۱۲	۲۲۷/۵۴۲	۲۲۶/۸۳۴	۳۰	۲۳۹/۶۹۳۰	۲۴۰/۳۲۴	۲۴۰/۲۸۵	۴۷	۲۴۸/۴۲۱	۲۴۸/۸۱۷	۲۴۸/۱۵۳		
۲	۰/۶۳۷۲	۰/۶۷۹۸	۰/۵۹۳۳	۵۸	۰/۸۱۶۴	۰/۵۸۹۳	۰/۶۲۴۴	۵۸	۰/۴۹۰	۰/۵۱۰	۰/۵۷۱		
۴	۰/۶۳۷۳	۰/۶۷۹۹	۰/۵۹۳۴	۱۹	۰/۸۱۷۰	۰/۵۸۹۶	۰/۶۲۴۸	۱۹	۰/۴۹۰	۰/۵۱۱	۰/۵۷۲		
۵۶	۰/۶۳۸۷	۰/۶۸۱۴	۰/۵۹۴۶	۵۷	۰/۸۱۷۸	۰/۵۸۹۹	۰/۶۲۵۴	۵۷	۰/۴۹۰	۰/۵۱۱	۰/۵۷۲	جهت باد	بهار
۵۳	۳۳۸/۴۵	۳۳۸/۹۹	۳۳۵/۵۱	۵۳	۳/۴۰۴۱	۳/۳۵۸۸	۳/۳۵۳۷	۵۳	۲/۰۶۲	۲/۰۵۹	۲/۰۸۴		
۵۲	۶/۲۶۶۴	۶/۱۹۴۶	۶/۲۰۳۳	۵۲	۶/۱۳۹۴	۶/۰۹۱۸	۶/۱۳۰۳	۵۲	۵/۷۶۴	۵/۷۲۶	۵/۷۸۴		
۷	۱۴/۱۷۷	۱۴/۱۸۱۶	۱۴/۱۲۱۶	۲۲	۱۳/۳۸۳۳	۱۳/۳۸۷۹	۱۳/۳۸۲۸	۴۱	۱۴/۰۴۱	۱۳/۹۹۴	۱۴/۰۲۳		
۱۳	۵۱/۶۴۴۳	۴۲/۶۹۸۰	۵۳/۱۴۹۵	۳۰	۵۴/۴۳۳۵	۵۴/۸۸۴۷	۵۴/۴۴۴۲	۴۷	۴۶/۰۶۷۲	۶۳/۶۳۴۶	۵۴/۶۴۴۷	جهت باد	بهار
۱۶	۴۹/۱۷۷۶	۴۰/۵۸۱۶	۵۰/۴۰۰۴	۲۸	۵۴/۱۱۷۶	۵۰/۸۱۰۶	۵۱/۰۵۰۴	۵۱	۴۴/۱۶۵۷	۶۲/۱۱۲۶	۵۲/۱۰۲۰		
۱۰	۴۸/۱۸۳۱	۳۹/۱۱۳۳	۴۸/۵۹۴۵	۳۷	۵۱/۲۱۱۲	۴۹/۳۶۰۲	۴۹/۴۷۵۵	۴۶	۴۲/۱۳۹۸	۵۸/۸۷۶۲	۴۹/۶۲۶۰		
۱۲	۴۷/۰۹۳۹	۳۸/۱۴۶۶	۴۷/۶۵۱۸	۲۹	۵۰/۶۷۵۲	۴۸/۸۳۵۳	۴۸/۹۲۵۵	۴۴	۴۰/۹۶۰۹	۵۸/۰۰۶۰	۴۹/۰۷۷۷		
۱۴	۴۶/۶۹۹۴	۳۸/۴۴۳۶	۴۷/۳۱۲۴	۳۱	۵۰/۲۴۴۵	۴۸/۲۵۱۳	۴۸/۴۰۰۳	۴۵	۴۰/۹۷۸۸	۵۷/۷۳۶۴	۴۸/۸۷۷۷	سرعت باد	بهار
۹	۴۶/۴۵۰	۳۸/۶۶۱۱	۴۷/۱۲۶۷	۲۶	۴۸/۰۲۱۹	۴۸/۰۲۴۶	۴۸/۰۳۴۶	۴۳	۴۰/۸۸۸۷	۵۷/۵۲۳۲	۴۸/۶۶۲۷		
۱۱	۴۶/۶۲۰۲	۳۸/۶۷۳۷	۴۷/۰۶۷۴	۵۶	۵۰/۳۳۹۵	۴۷/۸۱۶۵	۴۸/۲۳۰۵	۳۹	۴۰/۸۵۲۵	۵۷/۴۷۱۹	۴۸/۶۳۵۳		
۱۷	۴۶/۶۳۹۳	۳۸/۶۹۱۶	۴۷/۰۵۲۹	۲۸	۵۰/۲۳۸۱	۴۷/۷۷۲۳	۴۸/۱۶۶۴	۴۹	۴۰/۸۹۲۹	۵۷/۴۹۶۱	۴۸/۵۹۰۱		
۷	۷/۰۳۳۱	۷/۳۶۱۵	۵/۷۹۵۶	۲۲	۷/۰۳۶۶	۶/۲۰۷۵	۵/۹۲۴۴	۴۱	۵/۱۰۰۷	۶/۳۹۸۴	۵/۷۲۷۰	جهت باد	بهار
۳۲	۰/۸۵۳۲	۰/۷۶۶۹	۰/۶۶۶۹	۴۹	۰/۸۸۷۳	۰/۷۹۰۰	۰/۸۸۳۷	۴۹	۱/۰۳۱۶	۱/۰۳۱۶	۰/۶۸۲۷		
۱۵	۰/۵۱۹۲	۰/۹۲۵۷	۰/۶۸۸۳	۳۲	۰/۹۳۵۲	۰/۷۳۴۶	۰/۶۹۸۱	۳۲	۰/۵۰۸۹	۰/۹۶۸۰	۰/۵۹۳۵		
۴۹	۰/۴۸۸۰	۰/۸۹۷۱	۰/۶۶۲۱	۱۵	۰/۹۲۱۲	۰/۷۰۰۰	۰/۶۷۱۴	۱۵	۰/۴۷۴۹	۰/۹۴۳۷	۰/۵۶۳۴		
۱۷	۰/۴۲۲۱	۰/۸۹۰۱	۰/۶۵۲۲	۲۲	۰/۹۱۷۹	۰/۶۹۳۷	۰/۶۵۵۸	۳۹	۰/۴۶۸۹	۰/۹۳۸۶	۰/۵۵۱۳	سرعت باد	بهار
۵۶	۰/۴۷۳۱	۰/۸۹۹۷	۰/۶۴۵۵	۵۶	۰/۹۲۴۸	۰/۷۰۴۱	۰/۶۵۰۹	۱۷	۰/۴۶۳۹	۰/۹۴۱۰	۰/۵۷۱۷		
۵۴	۰/۴۳۳۵	۰/۸۹۸۷	۰/۶۴۲۴	۵۴	۰/۹۲۴۳	۰/۶۹۹۰	۰/۶۴۵۹	۷	۰/۴۶۵۷	۰/۹۳۸۸	۰/۵۴۳۲		
وزن ویژگی‌ها در معیار اطلاعات متقابل													
۱۵	۱/۲۹۱	۱/۲۶۰	۱/۲۶۴	۳۲	۱/۳۸۷	۱/۴۱۱	۱/۳۷۷	۵۲	۱/۴۰۸	۱/۶۴۶	۱/۶۲	جهت باد	اطلاعات متقابل
۱۴	۱/۱۹۱	۱/۱۷۸	۱/۱۶۸	۳۳	۱/۲۹۰	۱/۳۰۷	۱/۲۸	۴۹	۱/۳۴۳	۱/۴۲۵	۱/۴۱۹		
۱۶	۱/۱۴۰	۱/۱۳۵	۱/۱۲	۳۱	۱/۲۸۱	۱/۲۹۲	۱/۲۷۹	۴۸	۱/۳۰۸	۱/۳۵۳	۱/۳۴۲		
۱۳	۱/۱۳۰	۱/۱۱۸	۱/۱۱۴	۳۲	۱/۲۵۵	۱/۲۶۹	۱/۲۳۹	۵۰	۱/۲۸۹	۱/۳۲۴	۱/۳۰۷		
۱۷	۱/۱۲۰	۱/۱۰۶	۱/۱۱۰	۳۰	۱/۲۳۳	۱/۲۵۲	۱/۲۲۷	۵۱	۱/۲۷۷	۱/۳۰۹	۱/۲۹	سرعت باد	اطلاعات متقابل
۱۹	۱/۰۱۷	۰/۹۹۶	۱/۰۱۴	۳۶	۱/۱۶۴	۱/۱۷۰	۱/۱۶۳	۴۷	۱/۱۸۸	۱/۲۸۸	۱/۲۸۷		
۲۲	۰/۴۴۸	۰/۴۶۱	۰/۴۵۷	۳۹	۰/۴۶۷	۰/۴۶۶	۰/۴۵۸	۵۳	۰/۶۴۴	۱/۲۳۳	۱/۲۰۰		
۲۱	۰/۲۴۶	۰/۲۵۷	۰/۲۴۲	۳۸	۰/۳۱۰	۰/۳۰۷	۰/۳۰۱	۵۶	۰/۵۰۴	۰/۶۳۶	۰/۶۴۶		
۱۲	۰/۱۱۶	۰/۱۱۲	۰/۱۱۴	۲۹	۰/۱۱۰	۰/۱۱۱	۰/۱۰۹	۵۵	۰/۲۴۱	۰/۵۲۲	۰/۴۹۸	سرعت باد	اطلاعات متقابل
۱۱	۰/۱۰۶	۰/۱۰۸	۰/۱۰۶	۲۶	۰/۱۰۰	۰/۰۹۸	۰/۰۹۶	۴۴	۰/۲۱۷	۰/۲۴۷	۰/۲۳۳		
۹	۰/۱۰۰	۰/۰۹۸	۰/۱۰۰	۲۸	۰/۰۹۰	۰/۰۹۴	۰/۰۹۱	۴۱	۰/۲۱۶	۰/۲۲۱	۰/۲۱۱		
۸	۰/۰۹۹	۰/۰۹۷	۰/۱۰۰	۲۵	۰/۰۸۸	۰/۰۸۸	۰/۰۸۷	۴۵	۰/۲۰۶	۰/۲۲۰	۰/۲۰۶		
۶	۰/۰۹۵	۰/۰۹۵	۰/۰۹۵	۲۳	۰/۰۸۶	۰/۰۸۷	۰/۰۸۲	۴۲	۰/۲۰۶	۰/۲۱۰	۰/۱۹۸	سرعت باد	اطلاعات متقابل
۱۲	۰/۹۵۹	۰/۹۷۸	۰/۹۵۶	۲۹	۰/۹۷۵	۰/۹۷۵	۰/۹۷۷	۴۶	۱/۴۰۴	۱/۳۹۴	۱/۳۹۱		
۹	۰/۶۱۹	۰/۶۳۵	۰/۶۱۶	۲۶	۰/۷۱۲	۰/۷۰۴	۰/۷۰۵	۴۳	۱/۰۲۷	۱/۰۲۴	۱/۰۱۵		
۱۱	۰/۵۲۰	۰/۵۲۹	۰/۵۱۳	۲۸	۰/۶۳۲	۰/۶۱۸	۰/۶۲۴	۴۵	۰/۹۱۶	۰/۹۲۰	۰/۹۱۴		
۸	۰/۴۶۹	۰/۴۷۶	۰/۴۶۴	۲۵	۰/۵۵۵	۰/۵۴۴	۰/۵۵۰	۴۲	۰/۸۴۳	۰/۸۴۸	۰/۸۳۸	سرعت باد	اطلاعات متقابل
۶	۰/۴۳۵	۰/۴۳۸	۰/۴۲۷	۲۳	۰/۵۰۷	۰/۴۹۳	۰/۵۰۳	۴۰	۰/۸۰۴	۰/۸۰۴	۰/۷۹۵		
۲۲	۰/۳۳۷	۰/۳۳۵	۰/۳۳۲	۳۹	۰/۴۴۶	۰/۴۳۸	۰/۴۳۲	۴۱	۰/۶۹۴	۰/۶۹۶	۰/۶۸۶		
۷	۰/۳۲۵	۰/۳۳۴	۰/۳۲۳	۲۴	۰/۳۹۰	۰/۳۷۵	۰/۳۷۹	۴۴	۰/۶۵۰	۰/۶۵۳	۰/۶۴۲		
۱۰	۰/۲۷۷	۰/۲۹۱	۰/۲۷۳	۲۷	۰/۳۴۴	۰/۳۲۹	۰/۳۳۲	۵۵	۰/۵۱۴	۰/۵۰۷	۰/۴۹۸	سرعت باد	اطلاعات متقابل
۱۸	۰/۲۰۳	۰/۲۰۴	۰/۱۹۷	۳۵	۰/۳۲۵	۰/۳۱۷	۰/۳۱۵	۵۴	۰/۴۹۷	۰/۴۹۸	۰/۴۸۵		
۱۵	۰/۱۹۰	۰/۱۸۵	۰/۱۷۸	۳۱	۰/۲۸۶	۰/۲۸۴	۰/۲۷۷	۲۰	۰/۴۸۱	۰/۴۷۹	۰/۴۶۹		
۱۶	۰/۱۷۹	۰/۱۸۰	۰/۱۷۱	۳۲	۰/۲۸۱	۰/۲۷۸	۰/۲۷۵	۳۷	۰/۴۷۶	۰/۴۷۸	۰/۴۶۶		
۱۴	۰/۱۷۶	۰/۱۷۳	۰/۱۶۶	۳۴	۰/۲۷۶	۰/۲۶۷	۰/۲۶۴	۲	۰/۴۷۲	۰/۴۷۳	۰/۴۶۳	سرعت باد	اطلاعات متقابل
۱۹	۰/۱۷۶	۰/۱۷۱	۰/۱۶۴	۳۶	۰/۲۶۶	۰/۲۶۵	۰/۲۶۳	۵۹	۰/۴۶۹	۰/۴۵۹	۰/۴۵۳		

ادامه جدول ۴

مقادیر خطای جذر میانگین مربعات (RMSE) در انتخاب‌ویژگی												فصل	روش‌جستجو	کمیت
باند ۲۹			باند میانی			باند ۱۱			جهت باد	بهره باد	زمنستان			
ویژگی	اعتبارسنجی	آزمون	آموزش	ویژگی	اعتبارسنجی	آزمون	آموزش	ویژگی						
۹	۴۵/۲۵۵۹	۴۶/۳۵۱۹	۴۴/۹۲۵۵	۲۳	۴۵/۰۵۰۱	۴۶/۱۷۶۱	۴۶/۸۱۹۶	۵۳	۵۸/۵۷۷۸	۵۸/۸۴۱	۵۸/۹۲۶۰	جهت باد	بهره باد	زمنستان
۱۴	۴۵/۷۰۸۱	۴۶/۷۰۲۰	۴۵/۲۰۱۳	۳۱	۴۵/۶۳۷۷	۴۶/۷۹۱۴	۴۷/۴۸۹۲	۴۸	۵۹/۴۴۱۸	۵۹/۷۲۷	۵۹/۶۹۶۱			
۱۲	۴۶/۵۲۵۸	۴۷/۵۹۸۸	۴۶/۳۳۱۰	۲۹	۴۶/۱۷۵۷	۴۷/۴۷۵۲	۴۸/۱۵۴۹	۴۴	۶۰/۷۱۴۰	۶۰/۹۷۱	۶۱/۱۱۹۵			
۱۶	۴۸/۴۶۶۴	۴۹/۲۸۶۴	۴۸/۰۱۲۶	۲۷	۴۷/۸۳۸۷	۴۹/۴۳۴۱	۵۰/۱۱۱۵	۴۶	۶۳/۰۵۹	۶۳/۱۷۰	۶۳/۸۶۲۱			
۱۰	۴۹/۱۶۴۲	۵۰/۳۸۰۵	۴۸/۹۶۲۹	۵۲	۵۰/۳۸۳۰	۵۱/۴۷۱۳	۵۲/۰۷۷۲	۵۱	۶۴/۵۳۰	۶۴/۶۹۷	۶۵/۱۳۳۰			
۱۳	۳۳۲/۸۸۰۹	۳۳۲/۹۷۹۴	۳۳۲/۳۳۴۰	۳۰	۲۴۲/۸۴۲	۲۴۲/۸۷۹	۲۴۳/۰۴۵	۴۷	۲۴۳/۹۰۵	۲۴۳/۲۲۵	۲۴۴/۳۴۶			
۲۴	۰/۱۰۹۴	۰/۱۶۳۸	۰/۱۹۳۴	۴۶	۰/۱۶۷۳	۰/۱۹۳۴	۰/۱۶۱۰	۳۴	۰/۱۴۴۴	۰/۱۹۷۷	۰/۲۹۸۲	سرعت باد	بهره باد	زمنستان
۵۱	۰/۱۰۹۷	۰/۱۶۳۹	۰/۱۹۳۴	۵۱	۰/۱۶۷۵	۰/۱۹۳۵	۰/۱۶۱۲	۵۷	۰/۱۴۴۰	۰/۱۹۷۷	۰/۲۹۸۴			
۵۶	۰/۱۱۶۴	۰/۱۶۸۰	۰/۱۹۷۳	۵۶	۰/۱۷۰۹	۰/۱۹۶۵	۰/۱۶۹۰	۵۶	۰/۱۴۸۲	۰/۲۰۰۶	۰/۳۰۰۶			
۵۳	۲/۹۵۹۹	۲/۹۸۵۶	۲/۹۶۸۱	۵۳	۲/۹۲۸۴	۲/۸۷۱۲	۲/۹۰۰۰	۵۳	۱/۹۵۵۴	۱/۹۸۴۸	۱/۹۸۹۶			
۵۲	۶/۳۰۲۵	۶/۱۹۸۶	۶/۲۶۲۸	۵۲	۶/۲۷۲۷	۶/۱۴۹۲	۶/۲۴۳۳	۵۲	۵/۸۶۴۳	۵/۷۹۷۸	۵/۸۵۱۵			
۷	۱۲/۶۶۴۴	۱۲/۴۸۳۲	۱۲/۶۴۶۴	۲۲	۱۱/۵۷۲۰	۱۱/۴۴۴۸	۱۱/۵۵۷۲	۴۱	۱۱/۵۰۲۲	۱۱/۳۶۷۸	۱۱/۴۹۲۰			
۱۳	۵۲/۹۲۱۸	۵۲/۷۴۵۸	۴۷/۶۸۱۸	۳۰	۵۳/۶۴۹۵	۴۹/۴۹۶۸	۵۱/۲۵۷۱	۴۷	۶۹/۴۳۱۳	۵۷/۷۲۹۰	۶۵/۳۶۶۰	جهت باد	بهره باد	زمنستان
۱۰	۵۱/۹۳۹۶	۵۲/۰۵۵۹	۴۶/۶۶۷۸	۵۱	۵۱/۹۵۸۵	۴۹/۰۱۲۱	۴۸/۵۰۱	۵۱	۶۸/۰۸۴۸	۵۷/۱۸۴۰	۶۳/۸۶۵۱			
۱۶	۵۰/۶۶۲۵	۵۰/۰۶۹۷	۴۴/۹۳۳۰	۲۷	۵۰/۱۳۹۷	۴۷/۵۸۳۰	۴۶/۵۶۹۸	۴۴	۶۵/۷۷۲۳	۵۵/۱۷۳۴	۶۱/۳۹۰۹			
۱۲	۴۹/۱۳۹۸	۴۹/۶۸۸۲	۴۳/۸۵۸۳	۲۹	۴۹/۴۰۲۱	۴۷/۰۳۶۵	۴۶/۰۲۵۷	۴۶	۶۳/۶۹۵۴	۵۳/۰۸۱۴	۵۹/۹۳۲۱			
۱۴	۴۸/۷۲۵۴	۴۹/۱۰۱۱	۴۳/۶۴۰۳	۳۱	۴۸/۵۸۲۱	۴۶/۳۸۰۸	۴۵/۵۳۵۶	۴۸	۶۳/۳۶۶۶	۵۳/۲۷۸۰	۵۸/۸۸۵۵			
۹	۴۸/۵۳۷۷	۴۹/۰۲۴۴	۴۳/۴۸۶۷	۲۶	۴۸/۴۶۶۷	۴۶/۲۹۰۱	۴۵/۴۲۱۸	۴۳	۶۳/۲۷۷۲	۵۳/۲۱۹۲	۵۸/۸۳۴۹			
۱۱	۴۸/۴۱۵۰	۴۹/۰۸۷۶	۴۳/۴۴۵۲	۲۸	۴۸/۴۵۶۸	۴۶/۱۹۸۲	۴۵/۳۵۱۸	۵۳	۶۳/۲۹۰۳	۵۳/۳۹۷۹	۵۸/۸۰۲۲			
۷	۷/۲۳۹۳	۵/۶۵۷۰	۶/۱۴۹۸	۲۴	۶/۴۰۸۴	۶/۴۸۲۸	۶/۱۶۱۷	۴۱	۵/۰۵۲۶	۵/۵۳۹۳	۶/۰۸۰۵	سرعت باد	بهره باد	زمنستان
۴۹	۰/۵۶۵۸	۰/۵۳۸۵	۰/۴۲۳۴	۴۹	۰/۵۷۲۲	۰/۴۹۵۷	۰/۴۱۸۷	۴۹	۰/۷۱۸۱	۰/۴۳۶۹	۰/۴۶۸۸			
۳۲	۰/۳۹۹۰	۰/۴۵۱۳	۰/۳۲۰۷	۳۲	۰/۴۲۰۳	۰/۳۱۶۴	۰/۳۱۴۴	۳۲	۰/۶۰۶۵	۰/۳۲۶۱	۰/۳۷۸۲			
۱۵	۰/۳۴۶۳	۰/۴۴۴۳	۰/۲۸۳۹	۱۵	۰/۳۶۳۶	۰/۳۳۳۸	۰/۲۷۵۱	۱۵	۰/۵۶۹۹	۰/۲۸۴۸	۰/۳۴۵۶			
۵	۰/۳۲۷۲	۰/۴۱۱۸	۰/۲۷۲۳	۲۲	۰/۳۴۸۳	۰/۳۵۷۰	۰/۲۶۳۶	۳۹	۰/۵۵۸۷	۰/۲۷۱۷	۰/۳۳۴۸			
۱۶	۰/۳۲۲۶	۰/۴۱۴۸	۰/۲۶۸۱	۱۶	۰/۳۴۴۴	۰/۳۵۵۳	۰/۳۵۹۳	۱۷	۰/۵۵۴۱	۰/۲۶۱۹	۰/۳۳۴۸			
۱۷	۰/۳۱۷۳	۰/۴۱۱۲	۰/۲۶۳۷	۱۷	۰/۳۳۲۹	۰/۳۴۹۰	۰/۲۵۱۴	۱۶	۰/۵۴۶۸	۰/۲۶۰۳	۰/۳۳۰۱			
۵۴	۰/۳۱۵۰	۰/۴۰۹۷	۰/۲۶۱۲	۵۴	۰/۳۳۱۳	۰/۳۴۷۹	۰/۲۴۸۹	۵۴	۰/۵۴۵۶	۰/۲۵۶۶	۰/۳۲۷۴			
وزن ویژگی‌ها در معیار اطلاعات متقابل														
۱۸	۱/۲۶۶	۱/۲۷۰	۱/۲۷۱	۳۵	۱/۵۶۲	۱/۵۷۲	۱/۵۵۵	۵۲	۱/۶۶۴	۱/۶۷۳	۱/۶۶۵	جهت باد	اطلاعات متقابل	زمنستان
۱۵	۱/۰۷۷	۱/۰۷۲	۱/۰۸۶	۳۲	۱/۳۶۸	۱/۳۷۳	۱/۳۵۹	۴۹	۱/۴۶۳	۱/۴۷۶	۱/۴۷۵			
۱۴	۱/۰۰۶	۱/۰۱۳	۱/۰۲۱	۳۱	۱/۲۷۰	۱/۲۸۹	۱/۲۷۶	۴۸	۱/۳۹۷	۱/۴۰۳	۱/۴۰۵			
۱۷	۰/۹۹۶	۰/۹۹۸	۱/۰۰۲	۳۳	۱/۲۴۲	۱/۲۵۱	۱/۲۳۸	۵۰	۱/۳۵۵	۱/۳۷	۱/۳۶۷			
۱۳	۰/۹۴۰	۰/۹۳۹	۰/۹۵۵	۳۴	۱/۲۳۹	۱/۲۴۰	۱/۲۲۸	۴۷	۱/۳۵۲	۱/۳۴۹	۱/۳۵۱			
۱۶	۰/۹۳۲	۰/۹۳۷	۰/۹۴۳	۳۰	۱/۲۱۹	۱/۲۳۴	۱/۲۱۶	۵۱	۱/۳۴۲	۱/۳۴۶	۱/۳۵۰			
۱۹	۰/۸۵۷	۰/۸۶۳	۰/۸۵۴	۳۶	۱/۱۶۶	۱/۱۹۹	۱/۱۶۶	۵۳	۱/۳۰۷	۱/۲۹۸	۱/۲۹۷			
۲۲	۰/۳۵۷	۰/۳۴۷	۰/۳۵۴	۳۹	۰/۴۶۷	۰/۴۷۵	۰/۴۶۳	۵۶	۰/۴۹۰	۰/۴۹۲	۰/۴۸۶			
۲۱	۰/۱۶۵	۰/۱۶۶	۰/۱۷۰	۳۸	۰/۲۵۵	۰/۲۵۶	۰/۲۵۲	۵۵	۰/۴۴۲	۰/۴۳۲	۰/۴۴۰			
۱۲	۱/۰۵۴	۱/۰۲۳	۱/۰۳۶	۲۹	۱/۰۳۳	۱/۰۴۸	۱/۰۴۲	۴۶	۱/۲۸۵	۱/۲۵۴	۱/۲۷۸			
۹	۰/۷۱۵	۰/۶۸۵	۰/۶۹۳	۲۶	۰/۸۰۶	۰/۸۱۸	۰/۸۰۸	۴۳	۱/۰۱۹	۰/۹۸۴	۰/۹۹۶			
۱۱	۰/۶۱۳	۰/۵۷۵	۰/۵۸۵	۲۸	۰/۷۱۷	۰/۷۲۳	۰/۷۱۷	۴۵	۰/۹۲۶	۰/۸۹۶	۰/۹۰۹			
۸	۰/۵۷۷	۰/۵۴۴	۰/۵۵۱	۲۵	۰/۶۶۸	۰/۶۷۲	۰/۶۶۶	۴۲	۰/۸۷۱	۰/۸۳۷	۰/۸۵۲			
۶	۰/۵۴۲	۰/۵۰۹	۰/۵۱۹	۲۳	۰/۶۲۳	۰/۶۲۹	۰/۶۲۸	۴۰	۰/۸۳۴	۰/۸۰۶	۰/۸۱۵			
۷	۰/۴۵۳	۰/۴۲۰	۰/۴۲۱	۲۴	۰/۵۲۰	۰/۵۲۰	۰/۵۱۵	۴۱	۰/۷۲۸	۰/۶۹۸	۰/۷۱۲			
۲۲	۰/۴۴۱	۰/۴۱۸	۰/۴۲	۳۹	۰/۵۱۵	۰/۵۱۵	۰/۵۱۵	۴۲	۰/۶۷۲	۰/۶۴۶	۰/۶۵۹			
۱۰	۰/۳۹۸	۰/۳۶۶	۰/۳۷۱	۲۷	۰/۴۷۰	۰/۴۷۵	۰/۴۷۰	۵۴	۰/۴۷۶	۰/۴۵۷	۰/۴۶۵			
۱۸	۰/۲۱۸	۰/۲۱۱	۰/۲۰۴	۳۵	۰/۳۰۸	۰/۳۰۲	۰/۳۰۶	۲۰	۰/۴۶۸	۰/۴۳۹	۰/۴۵۴			
۱۵	۰/۲۰۲	۰/۱۹۴	۰/۱۸۹	۳۱	۰/۳۰۱	۰/۲۹۹	۰/۲۹۶	۲	۰/۴۵۰	۰/۴۳۴	۰/۴۳۸			



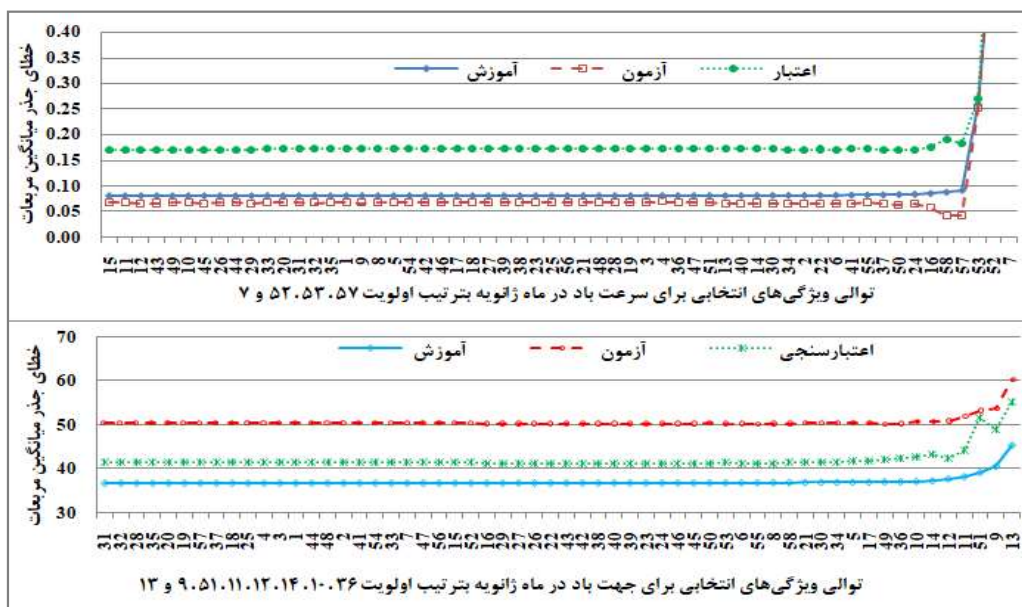
شکل ۳. ویژگی‌های انتخابی برای پیش‌بینی سرعت و جهت باد جستی در فصل‌های زمستان و بهار در باند ۲۹ با روش جستجوی پی‌درپی پیش‌رو.

۲-۳ انتخاب ویژگی با روش جستجوی پی‌درپی

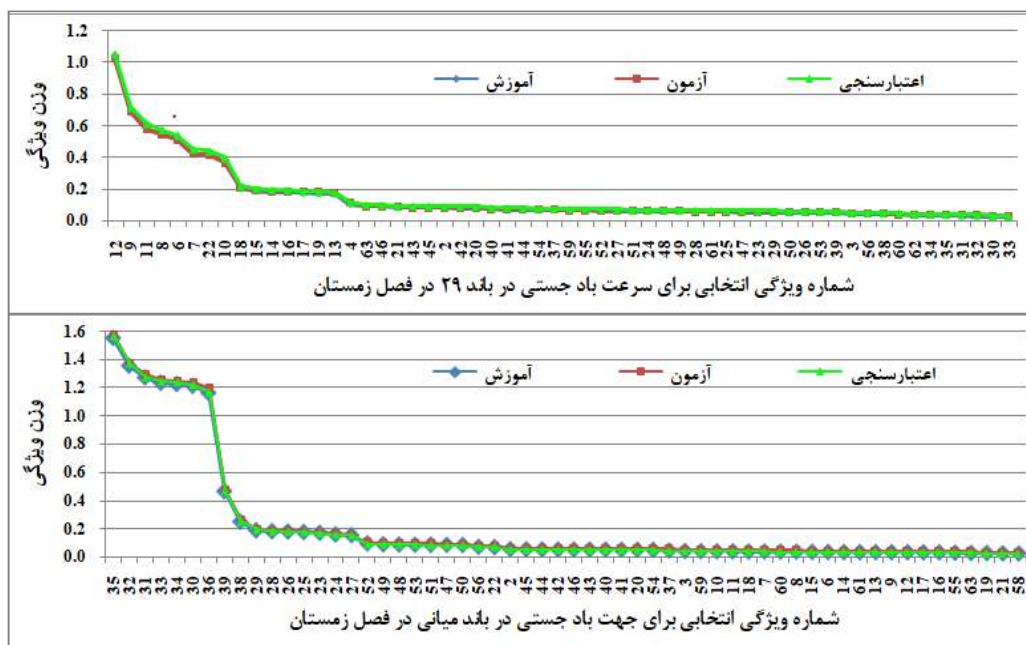
پسرو

افزایش کند. ویژگی‌های باقیمانده در مجموعه S را ویژگی‌های مطلوب در نظر می‌گیرند (وروریدیز و کاتروپلوس، ۲۰۰۵). نتایج اجرای این روش نیز برای فصل بهار و زمستان در جدول ۴ و به صورت نموداری در شکل ۴- الف برای ماه ژانویه در باند ۲۹ ارائه شده است.

ابتدا تمام ویژگی در مجموعه S قرار می‌گیرد. در هر عملیات تنها ویژگی‌ای حذف می‌شود که با حذف آن، مقدار RMSE شروع به کاهش کند. خاتمه عملیات زمانی است که با حذف یک ویژگی، RMSE مدل شروع به



(الف)



(ب)

شکل ۴. (الف) ویژگی‌های انتخابی با روش جستجوی پی‌درپی پسر و جهت و سرعت باد در ماه ژانویه در باند ۲۹ (ب) وزن ویژگی‌های انتخابی در فصل زمستان با روش اطلاعات متقابل برای سرعت باد جستی در باند ۲۹ و جهت باد جستی در باند میانی.

شکل ۴- ب برای سرعت باد جستی در باند شماره ۲۹ و برای جهت باد جستی در باند میانی ارائه شده است. وزن ویژگی در معیار MI در محور قائم شکل و ویژگی‌های معرفی شده در جدول ۲ نیز در محور افقی شکل قرار داده شدند. در این روش ویژگی‌هایی که وزن بیشتری را به خود اختصاص دادند، حداکثر ارتباط را با کمیت پیش‌بینی شونده دارند.

ویژگی‌های انتخابی ارائه شده در جدول ۴ برای جهت و سرعت باد جستی در باندهای فرودگاه مهرآباد با استفاده از RMSE به طور خلاصه در جدول ۵ آورده شده است.

۴ اجرای مدل رگرسیون

معادله رگرسیون خطی چندگانه به صورت زیر است:

(۳)

$$f(x_1 + x_2 + \dots + x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$$

که x_i متغیر مستقل است. ضرایب شیب خط (a_i) و عرض از مبدأ (b) ثابت هستند و معمولاً برای حل آنها از روش حداقل مربعات استفاده می‌شود. اگر رابطه فوق به شکل ضرب برداری برای باد جستی به کار رود، آنگاه رابطه کلی برای پیش‌بینی جهت و سرعت باد به شکل زیر خواهد بود:

$$\text{Gust Wind} = A + \sum_i^n \bar{F}_i \cdot \bar{M}_i \quad (۴)$$

که A مقدار ثابت و \bar{F}_i برداری است که در بردارنده شماره ویژگی‌هایی است که مشخصات آنها در جدول ۲ ارائه شده است. \bar{M}_i که به آن پیش‌بینی کننده هم گفته می‌شود، ضرایبی هستند که باید در مقادیر بردار ویژگی‌ها ضرب داخلی شوند.

۳-۳ معیار اطلاعات متقابل (Mutual Information, MI)

این معیار کمیتی است که وابستگی متغیرها را به یکدیگر بیان می‌کند و در قیاس با معیار تحلیل همبستگی که فقط روابط خطی موجود بین متغیرهای تصادفی را آشکار می‌سازد، توانایی آشکارسازی روابط خطی و غیرخطی موجود بین متغیرها را دارد (استیر و همکاران، ۲۰۰۵) و بنابراین برای سامانه‌های غیرخطی با دینامیک پیچیده چندبعدی نظیر پدیده‌های جوئی مثل طوفان بسیار مناسب است. این کمیت برای دو متغیر تصادفی گسسته مانند x و y به صورت زیر تعریف می‌شود:

$$MI(x, y) = \sum_{y \in Y} \sum_{x \in X} P(x, y) \log \left(\frac{P(x, y)}{P_1(x)P_2(y)} \right) \quad (۲)$$

که $P(x, y)$ تابع توزیع احتمال مشترک x و y و $P_1(x)$ و $P_2(y)$ به ترتیب توابع توزیع احتمال حاشیه‌ای x و y هستند.

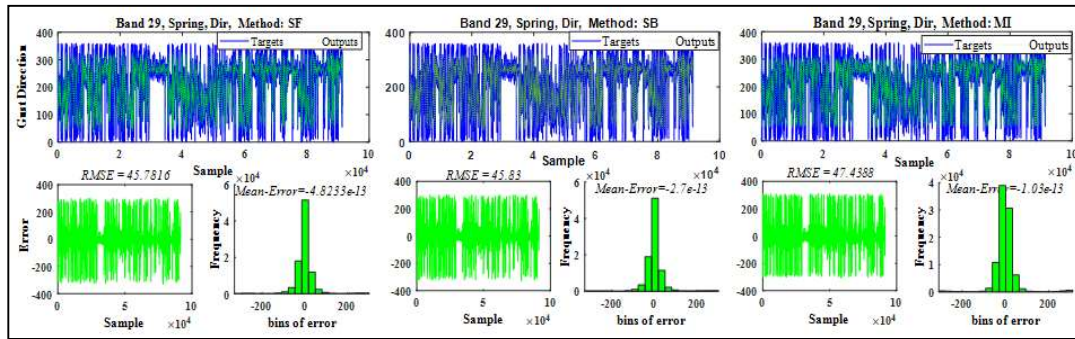
الگوریتم‌های متعددی برای انتخاب ویژگی بر مبنای معیار اطلاعات متقابل وجود دارد. در این تحقیق نیز همانند روش حیبی (۱۳۹۸) از روش حداقل افزونگی و حداکثر ارتباط استفاده شده است؛ یعنی اطلاعات متقابل بین ویژگی‌ها و برجسب‌های کلاس مرتبط با آن با استفاده از فضای ویژگی کوانتیده (quantized) تخمین زده می‌شود. به این منظور ابتدا ویژگی‌ها در ماتریس $N \times F$ قرار داده شدند که F تعداد ویژگی‌ها و N تعداد نمونه‌گیری‌ها است. البته داده‌های ویژگی نیز به دوازده تراز دسته‌بندی شدند. جهت باد از صفر تا ۳۶۰ درجه به هشت دسته تقسیم و برجسب زده شد. سرعت باد نیز به هفده گروه با فاصله ۱/۵ متر بر ثانیه تقسیم شد و سرعت‌های بیش از ۲۴ متر بر ثانیه نیز در یک دسته قرار داده شدند. نتایج این عملیات در جدول ۴ برای هر سه سنجنده و به صورت نمودار نیز در

جدول ۵. ویژگی‌های انتخابی برای جهت و سرعت باد جستی در باندهای فرودگاه مهرآباد در فصل زمستان و بهار. ویژگی انتخابی متفاوت در مقایسه دو فصل به رنگ قرمز و با زمینه زرد مشخص شده است.

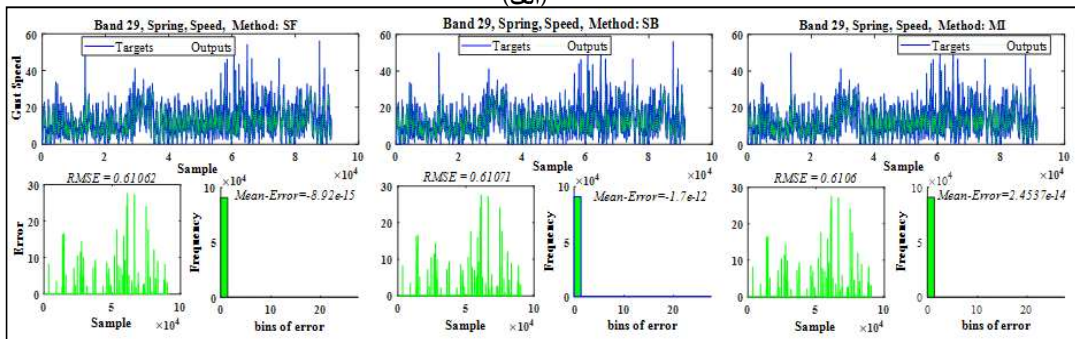
روش جستجو	فصل	جهت باد			سرعت باد		
		باندهای ۱۱	باندهای میانی	باندهای ۲۹	باندهای ۱۱	باندهای میانی	باندهای ۲۹
پسرو	بهار	۴۴	۲۹	۱۲	۵۳	۵۳	۵۳
	زمستان	۴۴	۲۹	۱۲	۵۲	۵۲	۵۲
پیشرو	بهار	۴۶	۲۷	۱۰	۴۹	۴۹	۴۹
	زمستان	۴۶	۲۷	۱۰	۴۹	۴۹	۴۹
اطلاعات متقابل	بهار	۴۷	۳۰	۱۳	۴۱	۴۱	۴۱
	زمستان	۴۷	۳۰	۱۳	۴۱	۴۱	۴۱

آبی در بالا نشان داده شده است. در پایین شکل، سمت چپ، اختلاف بین مقدارهای خروجی مدل با هدف و در سمت راست، بافت‌نگار فراوانی مقدارهای خطا به صورت دسته‌بندی شده نمایش داده شده است. البته در نمودار مربوط به بافت‌نگار، محور قائم از مرتبه ۱۰^۴ است. پهنای دسته‌ها برای جهت باد ۲۵ درجه و برای سرعت باد نیز یک متر برثانه انتخاب شده است.

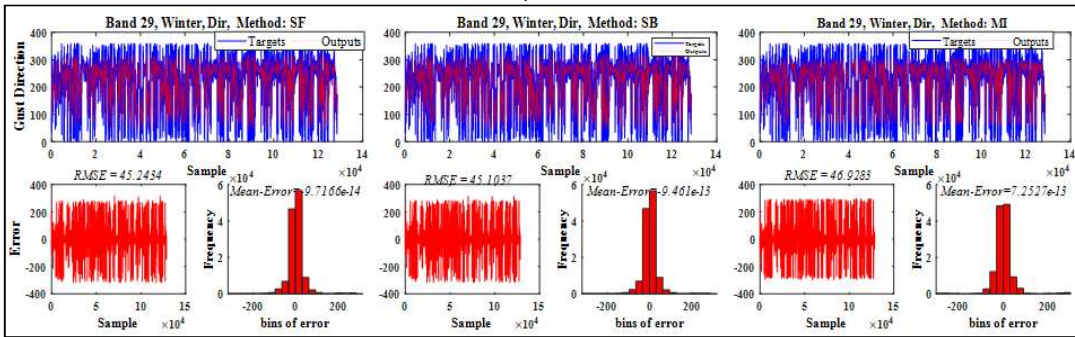
برای رسیدن به مدل رگرسیون مناسب برای فصل‌های زمستان و بهار، ویژگی‌های انتخابی با هر سه روش، یک‌به‌یک در رابطه (۴) قرار داده شدند و برای هر یک از آنها بردار ضرایب (\vec{M}_1) برای پیش‌بینی جهت و سرعت باد جستی به دست آمده است که نتایج آن در جدول ۶ و شکل‌های ۵ تا ۸ ارائه شده است. در شکل‌های ۵ و ۶ مقدار خروجی مدل با خط چین قرمز و مقدار هدف با خط



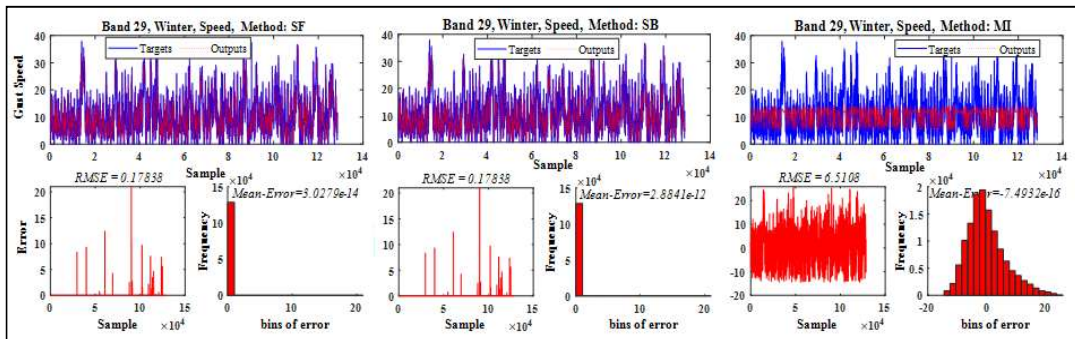
(الف)



(ب)

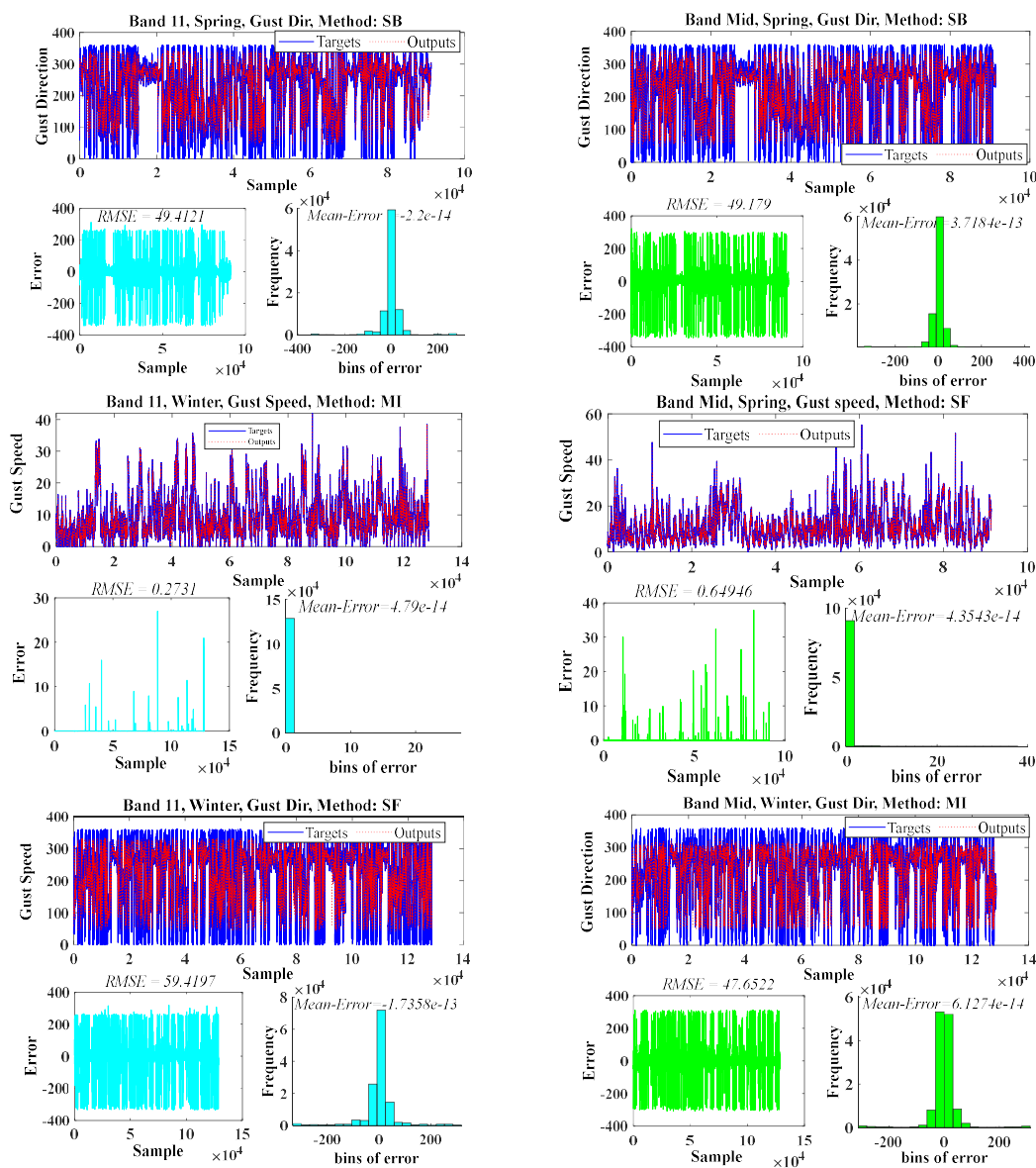


(ج)



(د)

شکل ۵. خروجی پیش‌بینی جهت و سرعت باد جستی برای باند شماره ۲۹ با معادله رگرسیون خطی و ویژگی‌های انتخابی از سه روش اطلاعات متقابل (ستون راست)، پسر (ستون وسط) و پیشرو (ستون چپ) در فصل بهار (رنگ سبز) و زمستان (رنگ قرمز). (الف) جهت باد در بهار (ب) سرعت باد در بهار (ج) جهت باد در زمستان (د) سرعت باد در زمستان.



شکل ۶. نمونه‌هایی از خروجی پیش‌بینی جهت و سرعت باد جستی با معادله رگرسیون خطی و ویژگی‌های انتخابی از سه روش. شکل‌های حاوی رنگ سبز متعلق به باند میانی و بقیه مربوط به باند ۱۱ است.

۵ بررسی نتایج

روش پسر (برای مثال در باند ۲۹) نشان می‌دهد که با حذف چهار ویژگی نمایش داده شده با اعداد ۱۰، ۱۲، ۱۰، ۵۲ و ۱۳ (جدول ۲) مقدار RMSE در اعتبارسنجی و آزمون به یک‌باره افزایش زیادی می‌یابد؛ بنابراین این ویژگی‌ها برای جهت باد انتخاب می‌شوند. در روش پیشرو، با

خطای جذر میانگین مربعات (RMSE) برای انتخاب ویژگی مناسب برای پیش‌بینی جهت و سرعت باد جستی در باندهای فرودگاه مهرآباد در جدول ۴ گنجانده شده است. بررسی انتخاب ویژگی مناسب برای جهت باد در

متوسط جهت باد در ده دقیقه هستند که سنجنده مستقر در باند اندازه‌گیری کرده است. البته متغیر متوسط جهت باد در باند در طی ده دقیقه در تمام حالت‌ها تأثیرگذار است. متغیرهای ۴۷ و ۵۱ (مؤلفه‌های سرعت باد در راستای باند پروازی) در باند ۱۱ در تمام روش‌ها انتخاب شدند که دلالت بر اهمیت این متغیرها دارد.

بررسی ویژگی‌های انتخاب‌شده سرعت باد در هر سه باند نشان می‌دهد ویژگی‌های انتخابی برای زمستان و بهار در روش پسر و مشابه هستند و می‌توان با متغیرهای فشار ایستگاه نسبت به سطح دریا، فشار لحظه‌ای و حداکثر سرعت باد باند در ده دقیقه، آنها را پیش‌بینی کرد. در روش پیشرو، ویژگی‌های زمستان، زیرمجموعه‌ای از ویژگی‌های بهار است و با حداکثر سرعت باد باند در ده دقیقه و انحراف جهت باد در ده دقیقه گذشته در باندهای ۱۱ و میانی ارتباط دارد. در اطلاعات متقابل، زمستان، تقریباً مشابه بهار است و با متغیرهای جهت باد لحظه‌ای، حداقل و حداکثر جهت باد در دو دقیقه، حداقل جهت باد در ده دقیقه و متوسط سرعت باد در ده دقیقه ارتباط دارد. در مرحله بعد، ویژگی‌های انتخابی با استفاده از رابطه (۴) برای تعیین ضرایب معادله رگرسیون به کار رفتند (جدول ۶). در این جدول متغیرهایی که بزرگ‌ترین ضریب را دارند و به عبارتی، بیشترین تأثیر مثبت را در پیش‌بینی سرعت و جهت باد دارند، با پس زمینه آبی مشخص شده است و متغیرهایی که اثر کاهش‌دهنده دارند، به رنگ قرمز و با زمینه زرد نشان داده شده‌اند.

مقایسه نتایج مطالعه کنونی با تحقیق حبیبی (۱۳۹۸)، که از انتخاب ویژگی و شبکه عصبی پرسپترون چندلایه برای پیش‌بینی باد جستی استفاده کرده است، نشان می‌دهد ویژگی‌های انتخابی از روش اطلاعات متقابل تقریباً در هر دو یکسان هستند، اما مقایسه ویژگی‌های انتخابی برای سرعت باد از روش جستجوی پی‌دی پی پیشرو با پیشروی شناور (حبیبی، ۱۳۹۸) نشان می‌دهد ویژگی‌های انتخابی از

انتخاب ویژگی‌های مناسب، RMSE نسبت به انتخاب قبلی کاهش دارد تا جایی که این کاهش متوقف می‌شود یا روند افزایشی می‌یابد. انتخاب ویژگی با معیار اطلاعات متقابل بستگی به وزن اختصاص‌یافته به ویژگی دارد؛ به عبارتی، ویژگی‌هایی که وزن بیشتری را به خود اختصاص داده‌اند، حداکثر ارتباط را با کمیت پیش‌بینی‌شونده دارند. در این جدول، ویژگی‌های انتخابی مناسب برای هر روش، پررنگ و با پس زمینه زرد مشخص شده است. تفسیر ویژگی‌های انتخابی از هر روش و برای هر باند برای سهولت مقایسه، در زیر نمودارها آورده شده است.

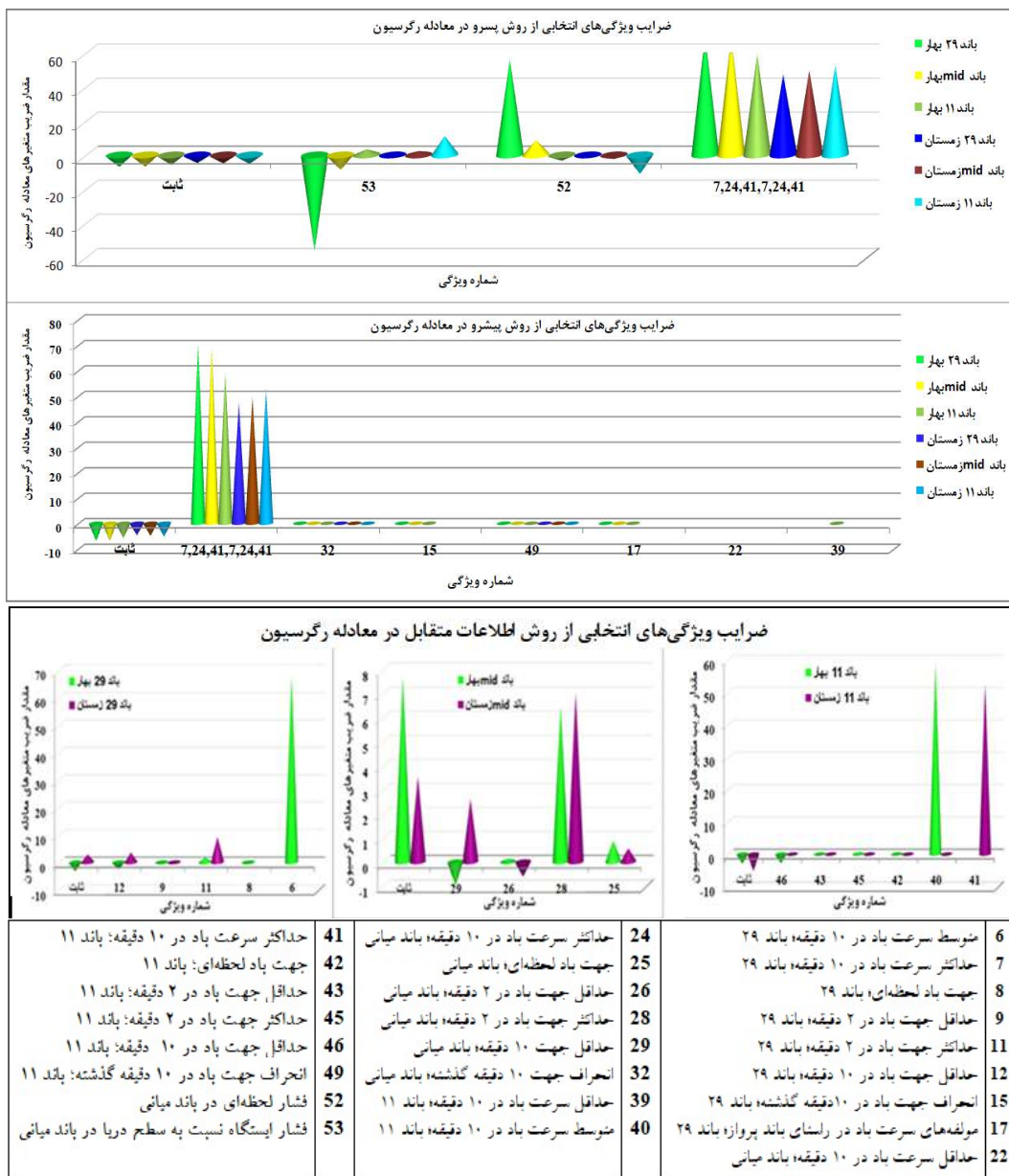
مقایسه ویژگی‌های انتخاب‌شده (جدول ۵) در روش پسر و برای سرعت باد در دو فصل نشان می‌دهد این ویژگی‌ها در هر سه باند مشابه هستند، اما برای جهت باد اندکی تفاوت وجود دارد. در روش پیشرو برای سرعت و جهت باد، ویژگی‌های زمستانی تعداد کمتری دارند، ولی زیرمجموعه‌ای از ویژگی‌های بهاری هستند. در روش اطلاعات متقابل نیز ویژگی‌ها مشابه هستند، اما همان‌گونه که مشاهده می‌شود، ترتیب انتخاب یکسان نیست؛ به عبارتی وزن‌های یکسانی ندارند. در این جدول، ویژگی‌های انتخابی متفاوت بین دو فصل با زمینه زرد رنگ آمیزی شده است.

مقایسه ویژگی‌های انتخابی روش‌ها برای جهت باد در هر سه باند نشان می‌دهد در روش‌های پسر و پیشرو، تقریباً ویژگی‌های یکسانی انتخاب شده‌اند، ولی در روش اطلاعات متقابل این‌گونه نیست. در ضمن در باندهای میانی و ۲۹، تعداد ویژگی‌های انتخابی در روش پیشرو، بیشتر از روش پسر است. نکته جالب اینکه متغیرهای برجسته‌شده با اعداد ۴۴، ۴۶ و ۴۷ در باند ۱۱، متغیرهای ۲۷، ۲۹ و ۳۰ در باند میانی و متغیرهای ۱۰، ۱۲ و ۱۳ در باند ۲۹ در روش‌های پسر و پیشرو دیده می‌شوند. این متغیرها به ترتیب متوسط جهت باد در دو دقیقه و حداقل و

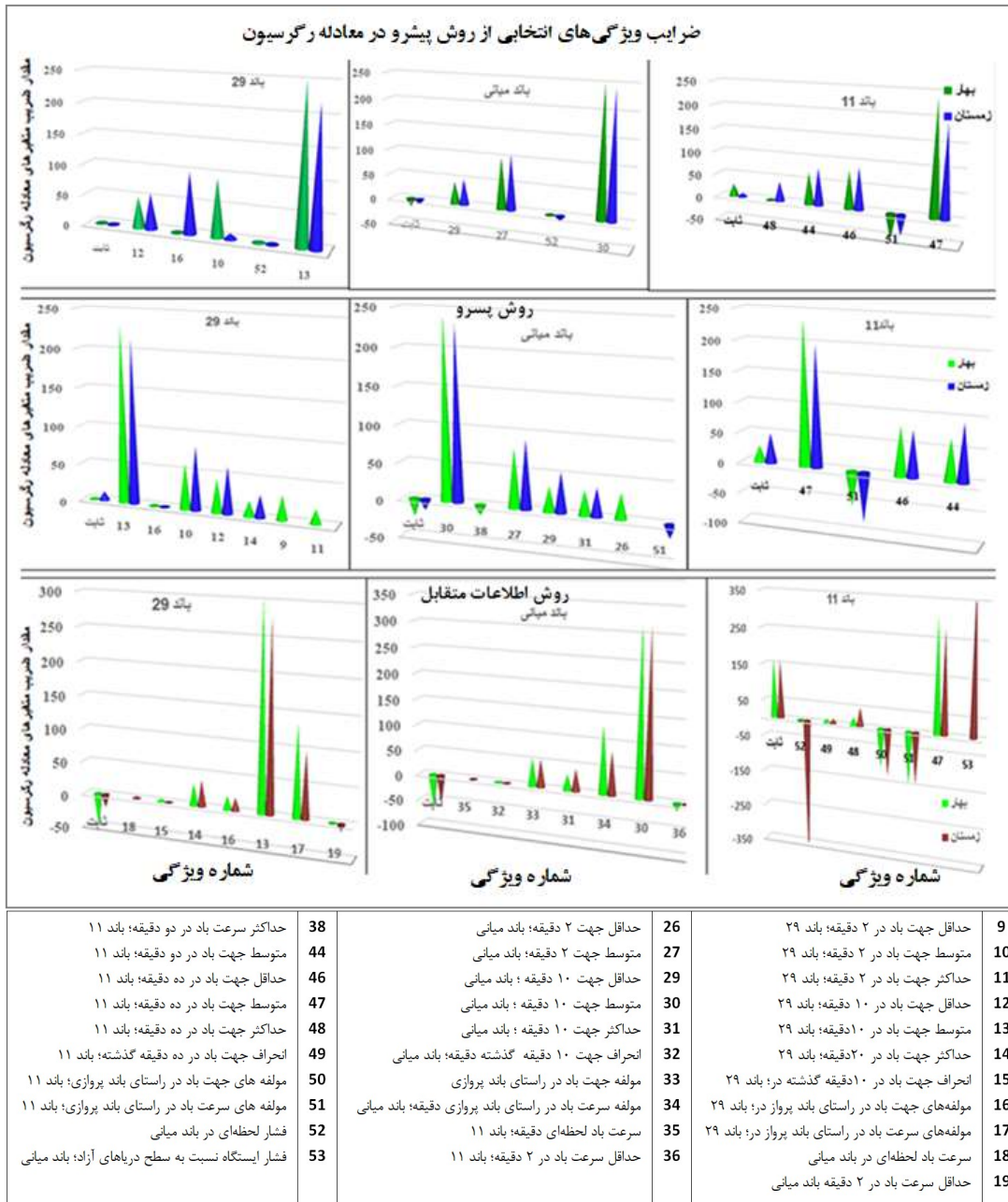
روش پیشرو، زیرمجموعه‌ای از روش پیشروی شناور هستند. البته ذکر این نکته در اینجا لازم است که روش پیشروی شناور از دقت بیشتری نسبت به روش پیشروی ساده برخوردار است؛ چون در آن امکان حذف ویژگی انتخابی در مراحل بعدی وجود دارد، اما در روش پیشروی این‌طور نیست. مزیت این روش نسبت به روش قبلی، سهولت محاسبات آن و نیاز نداشتن به تنظیمات شبکه عصبی پرسپترون چندلایه است.

جدول ۶. ضرایب ویژگی‌ها در معادله‌های رگرسیون به‌دست آمده برای جهت و سرعت باد جستی در باندهای فرودگاه مهرآباد در فصل زمستان و بهار. متغیرهایی که بیشترین تأثیر مثبت را دارند، با زمینه آبی مشخص شده‌اند. متغیرهایی که بیشترین تأثیر منفی را در پیش‌بینی دارند، با زمینه زرد و به رنگ قرمز مشخص شده‌اند.

روش جستجو	فصل	جهت باد				سرعت باد			
		بانند ۱۱	بانند میانی	بانند ۲۹	بانند ۱۱	بانند میانی	بانند ۲۹	بانند ۱۱	بانند ۲۹
پیشرو	بهار	ثابت	۲۸/۰۳۶	ثابت	-۱۶/۳۹۸	ثابت	-۷/۰۰۶	ثابت	-۷/۰۰۶
	زمستان	ثابت	۲۸/۰۳۶	ثابت	-۱۶/۳۹۸	ثابت	-۷/۰۰۶	ثابت	-۷/۰۰۶
اطلاعات متقابل	بهار	ثابت	۲۸/۰۳۶	ثابت	-۱۶/۳۹۸	ثابت	-۷/۰۰۶	ثابت	-۷/۰۰۶
	زمستان	ثابت	۲۸/۰۳۶	ثابت	-۱۶/۳۹۸	ثابت	-۷/۰۰۶	ثابت	-۷/۰۰۶



شکل ۷. مقادیر ضرایب متغیرهای به‌کاررفته در معادله رگرسیون برای پیش‌بینی سرعت باد جستی در باندهای فرودگاه مهرآباد. تفسیر شماره ویژگی، در جدول پایین آورده شده است.



شکل ۸. مقادیر ضرایب متغیرهای به‌کاررفته در معادله رگرسیون برای پیش‌بینی جهت باد جستی در باندهای فرودگاه مهرآباد. بهار با رنگ سبز و زمستان با رنگ آبی و بنفش نشان داده شده است. تفسیر شماره ویژگی، در جدول پایین آورده شده است.

پیش‌بینی شده با سرعت و جهت باد جستی ثبت شده در سنجنده مورد آزمایش مقایسه شد. نتایج بررسی بافت‌نگارهای فراوانی نشان می‌دهد که در کل،

۶ نتیجه‌گیری
در مرحله آخر با استفاده از ضرایب به‌دست‌آمده برای معادله خطی رگرسیون، سرعت و جهت باد جستی

در راستای باند پرواز در باند ۲۹، X_{52} فشار لحظه‌ای ایستگاه و X_{53} فشار ایستگاه نسبت به سطح دریا است.

۲-۶ معادلات پیشنهادی برای باند میانی

(۹)

$$DDmid_{Spring} = -24.93 + 241.94 X_{30} - 14.23X_{38} + 79.1X_{27} + 33.69X_{29} + 34.02 X_{31} + 34.5 X_{26} \quad (10)$$

$$DDmid_{Winter} = -15.09 + 233.33X_{30} - 18.14X_{51} + 91.36X_{27} + 54.14X_{29} + 38.56 X_{31} \quad (11)$$

$$FFmid_{Spring} = -6.91 + 69.29X_{24} + 0.006X_{49} + 0.028X_{32} + 0.011X_{15} - 0.242X_{22} \quad (12)$$

$$FFmid_{Winter} = -4.958 + 49.58 X_{24} + 0.001X_{49} + 0.007X_{32}$$

که X_{15} و X_{49} انحراف جهت باد در ده دقیقه گذشته در باندهای ۲۹ و ۱۱، X_{22} و X_{24} حداقل و حداکثر سرعت باد در ده دقیقه گذشته در باند میانی، X_{26} و X_{27} حداقل و متوسط جهت باد در دو دقیقه در باند میانی، X_{29} و X_{30} به ترتیب حداقل، متوسط و حداکثر جهت باد در ده دقیقه گذشته در باند میانی، X_{32} انحراف جهت باد در ده دقیقه گذشته در باند میانی، X_{38} حداکثر سرعت باد در دو دقیقه در باند ۱۱ و X_{51} مؤلفه‌های سرعت باد در راستای باند پروازی در باند ۱۱ است.

۳-۶ معادلات پیشنهادی برای باند ۱۱

(۱۳)

$$DD11_{Spring} = 28.036 + 237.5X_{47} - 57.56X_{51} + 80.97X_{46} + 69.34 \quad (14)$$

$$DD11_{Winter} = 50.182 + 198.42X_{47} - 82.46X_{51} + 76.14X_{46} + 94.81 X_{44}$$

ویژگی‌های انتخابی از روش‌ها در پیش‌بینی سرعت باد، خطای کمتری نسبت به پیش‌بینی جهت باد دارند.

ویژگی‌های انتخابی روش پیشرو جهت و سرعت باد جستی در باند میانی در قیاس با دو روش دیگر انحراف معیار کمینه دارد؛ بنابراین روابط حاصل از روش پیشرو، بهینه است. در باند ۱۱ ویژگی‌های انتخابی از روش پیشرو برای سرعت و ویژگی‌های انتخابی از روش پیشرو جهت پیش‌بینی جهت باد جستی بهینه هستند. در باند ۲۹ ویژگی‌های انتخابی از روش پیشرو جهت و جهت باد بهینه هستند.

با توجه به بررسی‌های انجام‌شده، روابط رگرسیون خطی برای پیش‌بینی جهت و سرعت باد جستی در فرودگاه خودکار مهرآباد برای فصل‌های زمستان و بهار به صورت زیر پیشنهاد می‌شود. البته در محاسبات از کمیت‌هایی که ضرایب کوچک‌تر از ۰/۰۱ دارند، می‌توان به دلیل کوچکی بیش از حد صرف نظر کرد.

۱-۶ معادلات پیشنهادی برای باند ۲۹

(۵)

$$DD29_{Spring} = -12.463 + 51.42X_{12} + 95.6X_{10} + 3.64X_{52} + 250.78X_{13} \quad (6)$$

$$DD29_{Winter} = -3.914 + 59.49X_{12} + 100.49X_{16} + 9.57X_{10} + 220.93X_{13} \quad (7)$$

$$FF29_{Spring} = -7.006 - 56.11X_{53} + 56.1X_{52} + 70.35X_7 \quad (8)$$

$$FF29_{Winter} = -4.74 + 0.16 X_{53} - 0.158X_{52} + 47.4X_7$$

که DD جهت باد، FF سرعت باد جستی، X_7 حداکثر سرعت باد در ده دقیقه در باند ۲۹، X_{10} متوسط جهت باد در دو دقیقه در باند ۲۹، X_{12} و X_{13} حداقل و متوسط جهت باد در ده دقیقه در باند ۲۹، X_{16} مؤلفه‌های جهت باد

سیزدهمین کنفرانس شبکه‌های توزیع نیروی برق، ۱۱

و ۱۲ اردیبهشت ۱۳۸۷-گیلان.

عرب عامری، م.، حبیبی، ف.، کلهر، ا.، ۱۳۹۳، پیش‌بینی

سرعت باد با استفاده از شبکه عصبی پرسپترون چند

لایه در فرودگاه مهرآباد: مجموعه مقالات شانزدهمین

کنفرانس ژئوفیزیک ایران، ۱۱۷-۱۲۲.

Agbo, G. A., Alfa, B., Ibeh, G. F., and Adamu, I. S., 2013, Application of regression and multiple correlation analysis to morning hours solar radiation in Lapai: International Journal of Physical Sciences, 8(27), 1437-1441, DOI: 10.5897/IJPS12.323, ISSN 1992-1950 © 2013 Academic Journals, <http://www.academicjournals.org/IJPS>.

Bluestein, H., 1992, Synoptic-Dynamic Meteorology in Midlatitudes, Volume I: Principles of Kinematics and Dynamics, Oxford University Press.

Fattahi, S., 2011, A comparative study of parametric and nonparametric regressions: Iranian Economic Review, 16(30), 19-43.

Gladysz, B., and Kuchta, D., 2008, Application of regression trees in the analysis of electricity load: Badania Operacyjne i Decyzje (Operations Research and Decisions), 4, 19-28.

Han, J., Kamber, M., and Pei, J., 2012, Data Mining: Concepts and Techniques, 3rd edition: Morgan Kaufmann Publishers, ISBN 978-0-12-3814791.

Hazra, A., and Gogtay, N., 2017, Biostatistics series module 10: Brief overview of multivariate methods: Indian Journal of Dermatology, 62(4), 358-366, doi: 10.4103/ijd.IJD_296_17.

Maksood, F. Z., and Achuthan, G., 2016, Analysis of data mining techniques and its applications: Journal of Computer Applications, 140(3), 6-14.

Moustris, K. P., Nastos, P. T., Larissi, I. K., and Paliatso, A. G., 2012, Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece: Advances in Meteorology, Article ID 894714, 8 pages, doi:10.1155/2012/894714.

Murphy, K., 2012, Machine Learning: A Probabilistic Perspective: The MIT Press

(۱۵)

$$FF11_{Spring} = -5.904 + 3.38X_{53} - 3.35X_{52} + 59.1X_{41}$$

(۱۶)

$$FF11_{Winter} = -5.248 + 11.13X_{53} - 11.14X_{52} + 52.52X_{41}$$

X_{44} , X_{41} , X_{47} و X_{46} به ترتیب حداکثر سرعت باد در ده

دقیقه، متوسط جهت باد در دو دقیقه، حداقل و متوسط

جهت باد در ده دقیقه در باند ۱۱ است. X_{51} مؤلفه‌های

سرعت باد در راستای باند پروازی در باند ۱۱، X_{52} فشار

لحظه‌ای ایستگاه و X_{53} فشار ایستگاه نسبت به سطح دریا

است.

منابع

حبیبی، آ.، ۱۳۹۵، آموزش کاربردی نرم‌افزار SPSS، ویرایش چهارم: پایگاه اینترنتی پارس‌مدیر، ۲۰۰ صفحه.

حبیبی، ف.، ۱۳۹۵، پیش‌بینی بارش ایستگاه بندرعباس با استفاده از روش رگرسیون لجستیک: مجموعه مقالات

هفدهمین کنفرانس ژئوفیزیک ایران، ۱۳-۱۷.

حبیبی، ف.، ۱۳۹۸، انتخاب ویژگی و پیش‌بینی باد گاستی

با شبکه عصبی پرسپترون چندلایه‌ای در ایستگاه

خودکار فرودگاهی: مجله ژئوفیزیک ایران، ۱۳(۳)،

۳۳-۵۲.

رضائی یوسفی، م.، ۱۳۸۶، انتخاب متغیرهای ورودی در

شناسایی سیستم‌ها و کاربرد آن در پیش‌بینی سری‌های

زمانی: پایان‌نامه کارشناسی ارشد در مهندسی برق-

گرایش کنترل، پردیس فنی دانشگاه تهران.

رضایی، م.، حسن‌زاده، م. ت.، قاسمی، س.، ۱۳۸۷،

استفاده از روش‌های رگرسیون خطی و رگرسیون

غیرخطی در تخمین روند تغییر جریان نشی مفره‌های

شبکه توزیع در مناطق آلوده سواحل جنوبی کشور:

- Cambridge, Massachusetts, ISBN 978-0-262-01802-9.
- Patlakas, P., Drakaki, E., Galanis, G., Spyrou, C., and Kallos, G., 2017, Wind gust estimation by combining a numerical weather prediction model and statistical post-processing: *Energy Procedia*, **125**, 190-198, www.elsevier.com/locate/procedia.
- Prasanna, V., Choi, H. W., Hong S. O. K., Kim, G. H., Lee, Y. G., and Kim, B. J., 2020, Surface wind gust prediction over Incheon international airport using the unified model: *Natural Hazards*, **103**, 1499–1535, DOI:10.1007/s11069-020-04047-z.
- Roundy, P. E., and Frank, W. M., 2004, Applications of a multiple linear regression model to the analysis of relationships between eastward- and westward-moving intraseasonal modes: *Journal of Atmospheric Science*, **61**(24), 3041-3048.
- Steuer, R., Daub, C. O., Selbig, J., and Kurths, J., 2005, Measuring distances between variables by mutual information, in Baier, D., Wernecke K. D. (eds), *Innovations in Classification, Data Science, and Information Systems. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg, https://doi.org/10.1007/3-540-26981-9_11.
- Vergara J. R., and Este´vez, P. A., 2014, A review of feature selection methods based on mutual information: *Neural Computing and Applications*, **24**(1), 175–186, DOI 10.1007/s00521-013-1368-0.
- Ververidis, D., and Kotropoulos, C., 2005, Sequential forward feature selection with low computational cost: 2005 13th European Signal Processing Conference, <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2005/defevent/papers/cr1411.pdf>.

Feasibility study of Gusty wind prediction using data mining and regression based on the sum of limited field data

Farideh Habibi ^{1*}

¹ Assistant Professor, Institute of Geophysics, University of Tehran, Tehran, Iran

(Received: 13 November 2020, Accepted: 31 January 2021)

Summary

This research has investigated the possibility of predicting the direction and speed of Gusty wind by using the information of Mehrabad Airport runway automatic station during the period of January 2013 to June 2013, the Metar report of Mehrabad Synoptic Station in the period of 2013, and regression method.

The data of the automatic station is taken from three sensors located in the band with a length of 4000 meters and a width of 45 meters, in southeast-northwest direction. The sensor number 29 and the sensor number 11, are at the northwestern end of the band and southeast edge of the band respectively. The Mid sensor location is at the middle of the band which distance from the band is 600 meters to the north direction.

First, all data (except the data of Gusty wind direction and speed measured by the sensors) was normalized to intervals 0.1-0.9. Second, all the data of sensors were randomly divided into three unequal parts: 70% of the data was stored for training, 50% of the remaining data was used for testing and the rest was used for validation. During the calculations, they were used instead of the original data. Third, the quantities were processed by using the three methods of feature selection: Sequential Forward Feature Selection(SFS); Backward(SBS) and Mutual Information(MI) with the method of the Maximum-Relevance and Minimum-Redundancy criterion. At this stage, selective features by every method were separately used in the linear regression method to predict the speed and direction of Gusty wind in the winter and spring seasons. The results were then compared with each other.

The results show that the selected features by SBS method for wind speed in winter are similar to spring, but their wind direction is slightly different. Selected features for winter Gusty wind with SFS method are a subset of the set of the selected features for spring. Selected features with MI are similar for the two seasons but with different weights.

The performance of the selected features for wind speed are better than for wind direction. The SFS method is optimal for selecting features of Gusty wind in the Mid runway. On runway 11, the SBS method and the SFS are optimized for predicting the Gusty wind speed and direction respectively. On runway 29, the SBF method is very suitable for selecting features related to Gusty wind speed and direction.

Finally, by examining the output of the models for each of the runways, an equation is provided to predict the direction and speed of the Gusty wind in each runway.

The quantity of predicted wind direction in runway 29 and 11 depends on the mean wind direction in 2 minutes, the minimum and mean direction in 10 minutes and the wind speed component along the runway, but the quantity of predicted wind speed depends on the maximum wind speed in 10 minutes on the runway, the instantaneous pressure of the station, and the pressure of the station relative to sea level.

The quantity of predicted wind direction in Mid runway depends on the minimum and mean wind direction in 2 minutes, minimum, mean and maximum wind direction in 10 minutes in runway Mid, maximum wind speed in 2 minutes and components of wind direction along with the runway 11. The quantity of predicted wind speed depends on the maximum and minimum wind speed in 10 minutes on the runway, deviation of wind direction during the last 10 minutes in the runways 29, Mid and 11.

Keywords: Backward Feature Selection (SFS), Forward Feature Selection (SBS), Gusty wind, Mehrabad Airport, prediction, regression, Mutual Information (MI)

*Corresponding author:

fhabibi@ut.ac.ir