

پیش‌بینی کوتاه‌مدت و بلندمدت بار الکتریکی با استفاده از روش‌های جدید یادگیری ماشین با در نظر گرفتن دما و زاویه ارتفاعی خورشید

رمضانعلی نقی‌زاده^۱

۱- استادیار- گروه مهندسی برق- دانشگاه صنعتی همدان- همدان- ایران

naghizadeh@hut.ac.ir

چکیده: هدف از این مقاله، به‌کارگیری و مقایسه عملکرد سه روش یادگیری ماشین شامل جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و اسپلاین‌های رگرسیون تطبیقی چندمتغیره (MARS) برای پیش‌بینی کوتاه‌مدت و بلندمدت بار الکتریکی است. اطلاعات ورودی به صورت بار الکتریکی ساعت قبلی، دمای هوا و زاویه فعلی ارتفاعی خورشید و روزهای تعطیل رسمی در نظر گرفته شده است. سه معیار مختلف برای مقایسه عملکرد شامل خطای جذر میانگین مربعات، خطای میانگین قدر مطلق و ضریب تعیین R^2 مورد استفاده قرار گرفته است. روش‌ها بر روی داده‌های ثبت شده واقعی مصرف بار الکتریکی یکی از پست‌های فوق توزیع شهر همدان و با استفاده از زبان برنامه‌نویسی متن‌باز R پیاده‌سازی شده است. داده دمای هوای منطقه از نزدیک‌ترین ایستگاه هواشناسی دریافت شده و زاویه تابش خورشید برای کل ساعات سال بر اساس موقعیت جغرافیای محل و روابط نجومی محاسبه شده است. نتایج نشان می‌دهد که روش‌های پیاده شده با دقت بسیار خوبی بار مصرفی را پیش‌بینی می‌نمایند و همچنین مدل RF در پیش‌بینی بار کوتاه‌مدت و روش SVM در پیش‌بینی بار بلندمدت عملکرد بسیار مناسبی ارائه می‌دهند و دقت بیشتری دارند.

واژه‌های کلیدی: پیش‌بینی بار کوتاه‌مدت، پیش‌بینی بار بلندمدت، یادگیری ماشین، جنگل تصادفی، ماشین بردار پشتیبان، اسپلاین‌های رگرسیون تطبیقی چندمتغیره

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.20.2.101

تاریخ ارسال مقاله: ۱۴۰۰/۰۶/۰۲

تاریخ پذیرش مشروط مقاله: ۱۴۰۰/۰۸/۲۹

تاریخ پذیرش مقاله: ۱۴۰۰/۰۹/۱۱

نام نویسنده‌ی مسئول: دکتر رمضانعلی نقی‌زاده

نشانی نویسنده‌ی مسئول: ایران - همدان - بلوار شهید فهمیده - خیابان مردم - دانشگاه صنعتی همدان، گروه مهندسی برق

۱- مقدمه

بهبود دقت پیش‌بینی بار الکتریکی به دلیل اهمیت بالای آن در عملکرد امن و اقتصادی سیستم قدرت در سالیان اخیر اهمیت بیشتری پیدا کرده است. بهبود دقت پیش‌بینی مصرف الکتریکی به دلیل تأثیرپذیری عوامل مختلف غیرقابل کنترل و غیرقطعی مانند تغییر آب و هوا، رشد اقتصادی، فعالیت‌های اجتماعی انسان‌ها، سیاست‌گذاری‌های دولتی و غیره چالش برانگیز است. از سالیان گذشته روش‌های متعددی برای پیش‌بینی بار الکتریکی مورد استفاده قرار گرفته است تا بهترین روش‌ها و راهکارها برای افزایش دقت پیش‌بینی شناسایی شده و یا توسعه داده شوند [۱] و [۲].

روش‌های ارائه شده برای پیش‌بینی بار الکتریکی را می‌توان به دو دسته‌بندی کلی روش‌های کلاسیک و روش‌های مبتنی بر هوش مصنوعی طبقه‌بندی نمود. برخی از نمونه‌های روش‌های کلاسیک مورد استفاده برای پیش‌بینی بار شامل مدل‌های رگرسیون [۳] و [۴]، هموارسازی نمایی [۵]، [۶]، [۷]، [۸]، [۹]، [۱۰]، [۱۱]، [۱۲]، [۱۳]، [۱۴]، [۱۵] و فیلتر کالمن [۱۶] است. پیاده‌سازی و کاربرد این روش‌ها ساده بوده و نیازمند بار محاسباتی زیادی نیست؛ اما خروجی این روش‌ها معمولاً دقت لازم و مناسب را ندارند. در سالیان اخیر روش‌های دقیق‌تر و کارآمدتری برای پیش‌بینی بار الکتریکی مانند شبکه‌های عصبی مصنوعی و یادگیری عمیق [۹] تا [۱۳]، منطق فازی [۱۴] و [۱۵] و سیستم‌های خبره [۱۶] مورد استفاده قرار گرفته است. با وجود افزایش دقت نسبت به روش‌های کلاسیک، این روش‌ها نیز معایب خود را دارند. برای نمونه انتخاب روش یادگیری، تنظیم مناسب تعداد لایه‌ها و یا تعداد نورون‌های شبکه عصبی، تنظیمات بهینه توابع عضویت و قواعد در منطق فازی جزء چالش‌های این روش‌ها به شمار می‌روند. همچنین با افزایش ورودی‌های کمکی به مسئله مانند دما، تابش و روزهای تعطیل، تنظیمات این روش‌ها بسیار پیچیده‌تر می‌شود. به‌هرحال انتخاب روش مناسب با توجه به مزایا و معایب هر کدام از آن‌ها همچنان نیاز به مطالعه و بررسی بیشتر دارد.

اخیراً استفاده از روش‌های جدید یادگیری ماشین برای مسائل پیش‌بینی توجه محققین را به خود جلب کرده است [۱۷]. یادگیری ماشین مبتنی بر درخت‌های تصمیم‌گیری (مانند RF)، ماشین بردار پشتیبان (SVM) و یادگیری ماشین مبتنی بر رگرسیون (مانند MARS) از جمله روش‌های جدید یادگیری ماشین است که با توجه به مزایای آن‌ها نسبت به شبکه‌های عصبی در سالیان اخیر در انواع کاربردهای مهندسی بیشتر مورد توجه قرار گرفته است. مدل RF یک مدل تک متغیره متشکل از ترکیب درختان رگرسیونی است و پارامترهای کمی برای تنظیم دارد. SVM یک مدل بسیار مؤثر و کارا در یادگیری ماشین است که توانایی حل مسائل غیرخطی با وجود داده یادگیری اندک را داشته و گزینه بسیار خوبی برای کاربردهای

پیش‌بینی است. MARS نیز یک ابزار جدید یادگیری ماشین است که دارای سرعت و انعطاف‌پذیری مناسب برای کاربردهای پیش‌بینی است. در این مقاله سه روش جدید و کارآمد یادگیری ماشین شامل RF، SVM و MARS برای پیش‌بینی بار الکتریکی مورد استفاده قرار گرفته است. نکته متمایز دیگر این است که در اکثر مقالات و مراجع قبلی، پیش‌بینی بار با روش مد نظر صرفاً برای یکی از حالت‌های کوتاه‌مدت یا بلندمدت مورد مطالعه قرار می‌گیرد؛ اما در این مقاله هر دو نوع پیش‌بینی کوتاه‌مدت و بلندمدت بار با هر سه روش پیاده‌سازی شده است که می‌تواند جهت انتخاب روش مناسب برای کاربردهای عملی راهگشا باشد. همچنین داده‌های استفاده شده برای این مطالعه شامل مقادیر اندازه‌گیری شده واقعی مصرف برق و دمای هوای یک منطقه مسکونی در شهر همدان بوده است. علاوه بر این زاویه ارتفاعی خورشید در هر ساعت در محل مورد مطالعه بر اساس روابط نجومی محاسبه و همراه با روزهای تعطیل رسمی به عنوان ورودی کمکی برای بهبود دقت پیش‌بینی استفاده شده است. در نهایت عملکرد روش‌ها ارزیابی و مقایسه شده است.

۲- تئوری روش‌ها

در این بخش تئوری مدل جنگل تصادفی، ماشین بردار پشتیبان و اسپلاین‌های رگرسیون تطبیقی چند متغیره با هدف پیش‌بینی سری زمانی شرح داده شده است. همچنین روابط مدل‌سازی موقعیت ساعتی خورشید در طول سال به عنوان یکی از ورودی‌های کمکی ارائه شده است.

۲-۱- جنگل تصادفی

جنگل تصادفی یک روش قدرتمند یادگیری گروهی است که توسط بریمان [۱۸] معرفی شده است. جنگل تصادفی را می‌توان به عنوان گروهی از درختان طبقه‌بندی و رگرسیون (CART) در نظر گرفت زیرا که در یک RF، مدل‌های CART چندگانه ایجاد شده به عنوان مدل‌های پایه استفاده می‌شوند. این روش برای کاربردهای طبقه‌بندی، مسائل رگرسیون و یادگیری بدون نظارت قابل استفاده است. این روش در زمینه‌های علمی زیادی مورد استفاده قرار گرفته است.

جنگل تصادفی مشتمل بر تعداد زیادی درخت تصمیم است که می‌توان از آن‌ها برای ایجاد یک مدل رگرسیون استفاده نمود. این درخت‌ها در زیر فضاهای منتخب تصادفی در فضای ویژگی مورد نظر رشد می‌کنند. اساس جنگل‌های تصادفی بر ترکیب مجموعه‌ای از درخت‌های تصمیم باینری است که هر کدام از آن‌ها بر اساس بهره‌گیری از نمونه‌برداری بوت استرپ^۸ از نمونه یادگیری و زیرمجموعه‌ای از ویژگی‌ها (مانند متغیرهای ورودی) که به صورت انتخاب تصادفی در هر گره ساخته می‌شوند، رشد درختان تا یک رسیدن به محدودیت حداکثری تعیین شده ادامه می‌یابد. رفتار

۲-۲- مدل ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) نیز یک روش یادگیری ماشین است که در مسائل رگرسیون به صورت گسترده استفاده شده است. SVM بر اساس حداقل سازی ریسک ساختاری است که خطای تجربی و نیز پیچیدگی را به طور همزمان مینیمم می‌کند [۲۱]. این ویژگی قابلیت تعمیم را برای SVM ایجاد می‌کند که باعث می‌شود این مدل برای کاربردهای پیش‌بینی بسیار مناسب باشد [۲۲]. مفهوم اساسی این روش این است که بردار ورودی فضای x با یک تابع کرنل مناسب غیرخطی که با عنوان $\phi(x)$ شناخته می‌شود، به یک فضای با ابعاد بالاتر نگاشت می‌شود. با استفاده از چنین تابع کرنل غیرخطی حل مسئله رگرسیون پیچیده غیرخطی فضای ورودی را تسهیل می‌کند. برای شرح خلاصه مدل SMV فرض کنید (x, y) یک مجموعه از متغیرهایی باشد که $x \in \mathbb{R}^m$ بردار ورودی با m عضو به عنوان پیش‌بین‌ها و y بردار خروجی باشد. تابع تخمین SVM برای هدف رگرسیون را می‌توان به صورت زیر بیان نمود:

$$y = f(x) = w \cdot \phi(x) + b \quad (1)$$

که در آن w یک بردار وزنی نشان‌دهنده ضریب رگرسیون و b یک عدد اسکالر و نشان‌دهنده بایاس است. این دو پارامتر را می‌توان با حداقل سازی تابع ریسک مرتب شده که به صورت زیر تعریف می‌شود تعیین نمود:

$$R_{SVMs}(C) = \frac{1}{2} \|w^2\| + C \frac{1}{n} \sum_{i=1}^n L(x_i, y_i) \quad (2)$$

که در آن C ضریب جریمه خطا، $L(x_i, y_i)$ تابع تلف و $\frac{1}{2} \|w^2\|$ عبارت منظم‌سازی است. C پارامتری است که دارای مقادیر مثبت است و بر اساس تنظیم خطا تعیین می‌شود. متغیرهای پشته ξ_i^* و ξ_i^* حدود انحراف اضافی بالا و پایین هستند. این دو متغیر خطاهای یادگیری را با تابع تلف جریمه می‌کنند. نهایتاً یک مسئله بهینه‌سازی بر اساس تابع هدف مرتبه دوم به صورت زیر تعریف می‌شود:

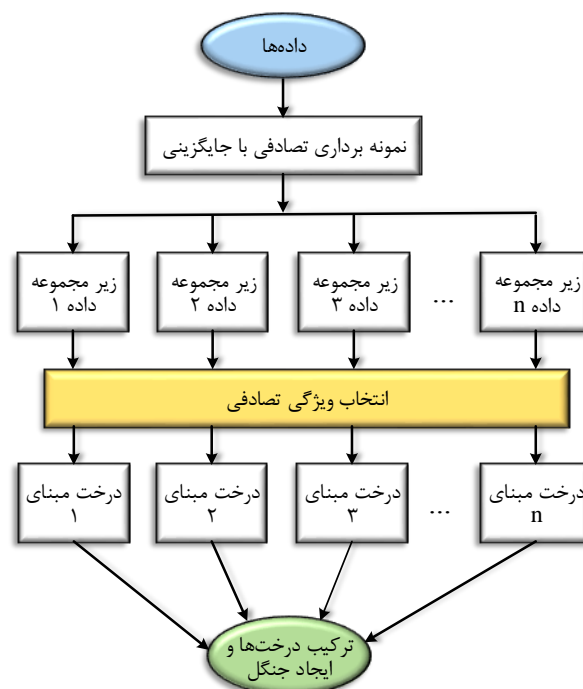
$$\text{Minimize } R_{SVMs}(w, \xi^{(*)}) = \frac{1}{2} \|w^2\| + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

$$\text{Subject to } \begin{cases} y_i - w\phi(x_i) - b_i \leq \varepsilon + \xi_i \\ w\phi(x_i) - b_i - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \end{cases} \quad (4)$$

که در آن ε نشان‌دهنده تلورانس خطا و l تعداد عناصر مجموعه یادگیری است. توابع کرنل مختلفی مانند چندجمله‌ای، پایه شعاعی

تصادفی با دو راهکار در این روش استفاده می‌شود. اولین مورد این است که در ایجاد درخت‌ها از زیرمجموعه‌های تصادفی از مشاهده‌ها (نمونه‌برداری بوت استرپ) استفاده می‌شود. دومین مورد این است که در ایجاد هر انشعاب یا شاخه در یک درخت از زیرمجموعه‌های تصادفی از متغیرهای کاندید استفاده می‌شود [۱۹].

طبیعت تصادفی چنین مجموعه‌ای از درخت‌ها نسبتاً ناپایدار است، زیرا یک تغییر کوچک در داده یادگیری می‌تواند منجر به تغییر شدید ساختار درخت گردد. این امر منجر به ایجاد تفاوت در پیش‌بینی‌های منفرد هر درخت خواهد شد. بر این اساس خطای OOB^۱ تعریف می‌شود که خطای ناشی از نقاط یادگیری است که در مجموعه‌های یادگیری بوت استرپ قرار ندارند. مزیت بزرگ جنگل تصادفی این است که می‌توان به صورت متوالی و متقاطع در حین رشد درخت‌ها اعتبار سنجی را انجام داد. یادگیری با پایدار شدن خطای OOB متوقف می‌شود. نکته جالب توجه دیگر این است که هر چقدر درجه تصادفی بودن در فرآیند یادگیری بیشتر باشد، منجر به نتایج دقیق‌تر در فرآیند یادگیری گروهی می‌شود. دو روش گام به گام شامل پیش رو و پس رو در مدل RF استفاده می‌شود. در گام پیش رو یک مدل با تطابق بیش از حد با تعداد زیاد گره‌های اضافی ایجاد می‌شود. به طور متقابل، در گام پس رو، یک فرآیند هرس برای زدودن گره‌های اضافی اعمال می‌گردد. در نتیجه در گام‌های پس رو با حذف متغیرهای غیرضروری دقت پیش‌بینی بهبود می‌یابد [۱۸]. فرآیند تشکیل جنگل تصادفی در شکل (۱) نشان داده شده است.



شکل (۱): فرآیند تشکیل جنگل تصادفی [۲۰]

که در این روابط β زاویه ارتفاعی خورشید نسبت به افق، H زاویه ساعتی، n شماره روز (در روز اول ژانویه $n = 1$ و در روز ۳۱ دسامبر $n = ۳۶۵$)، L زاویه عرض جغرافیایی در محل، و δ زاویه انحراف زمین است.

۲-۵- معیارهای ارزیابی عملکرد

در این مقاله از سه معیار خطای جذر میانگین مربعات ($RMSE$)^{۱۳}، خطای میانگین قدر مطلق (MAE)^{۱۴} و ضریب تعیین^{۱۵} (R^2) برای ارزیابی دقت پیش‌بینی RF، SVM و MARS استفاده شده است. این معیارها بر اساس روابط زیر محاسبه می‌شوند:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{\text{Observed}} - P_{\text{Predicted}})^2} \quad (۹)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_{\text{Observed}} - P_{\text{Predicted}}| \quad (۱۰)$$

$$R^2 = \frac{\sum_{i=1}^n (P_{\text{Predicted}} - P_{\text{Mean}})^2}{\sum_{i=1}^n (P_{\text{Observed}} - P_{\text{Mean}})^2} \quad (۱۱)$$

که در آن‌ها n تعداد مشاهدات از سری زمانی، P_{Observed} توان واقعی مشاهده شده، $P_{\text{Predicted}}$ توان پیش‌بینی شده، و P_{Mean} مقدار متوسط توان مشاهده شده است. $RMSE$ ملاک خوبی برای تعیین مقادیر خطای بزرگ و MAE نشان‌دهنده مقدار متوسط خطای پیش‌بینی است. لازم به ذکر است که شاخص $RMSE$ به دلیل استفاده از توان ۲ مقادیر خطاها منجر به حساسیت بیشتر این شاخص نسبت به مقادیر خطای بزرگ یا پیش‌بینی‌های پرت خواهد شد؛ درحالی‌که در شاخص MAE که بر اساس قدر مطلق خطا تعریف شده، مقادیر خطای بزرگ همانند شاخص $RMSE$ در آن نمود پیدا نمی‌کند. البته تفسیر و درک شاخص MAE نسبت به $RMSE$ ساده‌تر بوده و اثر مقادیر خطاها به یک نسبت در آن لحاظ می‌شود. درهرحال در این مقاله از هر دو شاخص استفاده شده و به صورت کلی، مقادیر کمتر $RMSE$ و MAE نشان‌دهنده دقت بیشتر پیش‌بینی است. شاخص سوم مورد استفاده نیز R^2 است که یک معیار تعیین‌کننده رابطه بین مقادیر مشاهده شده و پیش‌بینی شده بار الکتریکی است. مقادیر بالاتر R^2 نشان‌دهنده پیش‌بینی دقیق‌تر است و در حالت ایدئال R^2 برابر با ۱ خواهد شد.

۳- داده‌های مسئله، پیاده‌سازی روش‌ها و نتایج

در این بخش ابتدا کلیات مطالعه موردی و نحوه جمع‌آوری داده‌های سری‌های زمانی بار مصرفی، دما و زاویه ارتفاعی خورشید ارائه شده است. در ادامه نحوه پیاده‌سازی و تنظیم پارامترها و نتایج

گوسی تعمیم یافته ($GRBF$)^{۱۶} و پایه شعاعی نمایی برای نگاشت فضای ورودی به فضای ویژگی مرتبه بالا قابل استفاده است. در این مقاله از تابع کرنل GRBF به صورت $k(x_i, x) = \exp(-\gamma |x_i - x|^2)$ استفاده شده است. سه پارامتر C ، γ و ε باید به خوبی تنظیم شوند تا عملکرد مناسبی برای پیش‌بینی توسط این روش به دست آید.

۲-۳- اسپلاین‌های رگرسیون تطبیقی چند متغیره

اسپلاین‌های رگرسیون تطبیقی چند متغیره یا MARS یک مدل رگرسیون خطی گام به گام است که برای اهداف پیش‌بینی قابل استفاده است. مزیت اصلی این مدل این است که رابطه غیرخطی پیچیده بین پیش‌بین و متغیرهای پاسخ را به صورت ذاتی در برمی‌گیرد. این مدل دارای شکل کلی زیر است [۲۳]:

$$y = \beta_0 + \sum \beta_{ij} (x_i - \tau_j)^* + \sum_{i \neq k} \beta_{ijkl} (x_i - \tau_j)^* (x_k - \tau_l)^* + \dots \quad (۵)$$

که در آن β ضرایب رگرسیون هستند و $(x_i - \tau_j)^*$ یک تابع نگاشت بریده است که می‌تواند مثبت یا منفی باشد. تابع نگاشت فوق به صورت $y = \text{maximum}(0, c - x)$ و $y = \text{maximum}(0, x - c)$ در فضای پیش‌بینی (x) اعمال می‌شود تا فضای خروجی (y) را ایجاد کند. در این توابع c مقدار آستانه است که به عنوان گره نیز شناخته می‌شود. در معادله فوق امکان در نظر گرفتن تعامل بین تمام پیش‌بین‌ها با استفاده از عبارات $(x_i - \tau_j)^* (x_k - \tau_l)^*$ مقدور است. از یک تلاقی بین دو اسپلاین مجاور و یک گره برای حفظ پیوستگی توابع پایه استفاده می‌شود [۲۴].

۲-۴- موقعیت خورشید

موقعیت خورشید در میزان روشنایی طبیعی و نیز مقدار مصرف انرژی الکتریکی جهت روشنایی تأثیر دارد. موقعیت دقیق خورشید در هر زمان و هر مکان بر روی کره زمین قابل محاسبه بوده و بر اساس زاویه ارتفاعی خورشید^{۱۱} بیان می‌شود. این زاویه برای هر ساعت در کل سال و بسته به عرض جغرافیایی محل بر اساس روابط زیر قابل محاسبه است [۲۵]:

$$\beta(H, n) = \sin^{-1} [\cos L \cos \delta(n) \cos H + \sin L \sin \delta(n)] \quad (۶)$$

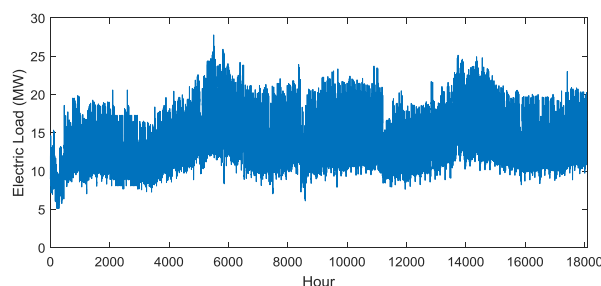
$$\delta(n) = 23.45 \sin \left[\frac{360}{365} (n - 81) \right] \quad (۷)$$

$$H(h) = 15^\circ \times (\text{hours before solar noon}) \quad (۸)$$

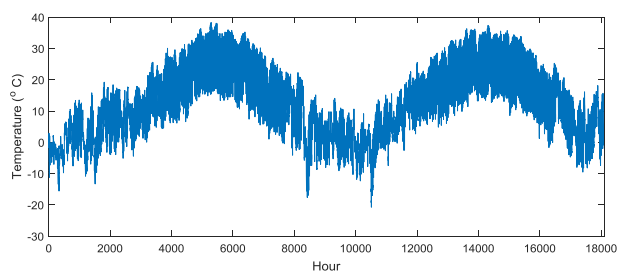
پیش‌بینی سه روش RF به همراه SVM و MARS ارائه و عملکرد و دقت آن‌ها مقایسه شده است.

۳-۱- داده‌های مسئله

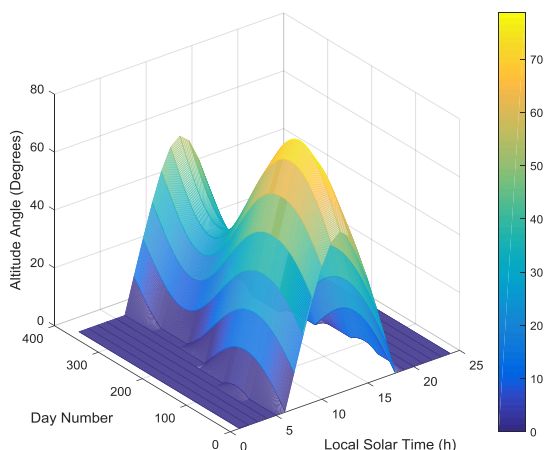
روش جنگل تصادفی به همراه SVM و MARS بر روی داده اندازه‌گیری شده واقعی پیاده‌سازی شده است. این داده مصرف یکی از پست‌های فوق توزیع ۶۳/۲۰ کیلوولت شهر همدان است که با کنتور SL7000 با کلاس دقت ۰/۲ به صورت ساعتی اندازه‌گیری شده است. فیدرهای خروجی این پست شامل مصارف تعرفه خانگی، عمومی و تجاری است. بازه زمانه اندازه‌گیری از تاریخ ۱۸ آذر ۱۳۹۴ تا ۹ دی ۱۳۹۶ بوده و شامل ۱۸۰۶۶ داده بار ساعتی است. شایان ذکر است که میانگین مصرف ۱۴/۲۹۸۴ مگاوات است. نمودار بار مصرفی در شکل (۲) نشان داده شده است. علاوه بر بار الکتریکی، داده‌های دمای ایستگاه هواشناسی نزدیک به پست (با فاصله ۴ کیلومتر از محل پست) نیز مطابق با بازه زمانی بار مصرفی ساعتی جمع‌آوری شده است. لازم به ذکر است که داده‌های دما به صورت ۱۰ دقیقه‌ای بوده و با میانگین‌گیری به صورت ساعتی تبدیل شده است. داده‌های ساعتی دما به عنوان یکی از ورودی‌های مدل‌های پیش‌بینی استفاده شده است. میانگین دمای محیط ۱۲/۴۲۱۹ درجه سانتی‌گراد بوده است. نمودار تغییرات دما در شکل (۳) نشان داده شده است. نمودار زاویه ارتفاعی خورشید نیز بر اساس ساعت و روز (بر اساس تقویم شمسی) در شکل (۴) ارائه شده است. لازم به ذکر است که اطلاعات ورودی به مدل‌های توسعه داده شده جهت پیش‌بینی به صورت بار الکتریکی ساعت قبلی، دمای هوا و زاویه فعلی ارتفاعی خورشید و روزهای تعطیل مطابق با تقویم رسمی کشور در نظر گرفته شده است.



شکل (۲): نمودار سری زمانی ساعتی مصرف الکتریکی مورد مطالعه



شکل (۳): نمودار سری زمانی ساعتی دمای هوا در منطقه مورد مطالعه



شکل (۴): نمودار زاویه ارتفاعی محاسبه شده خورشید

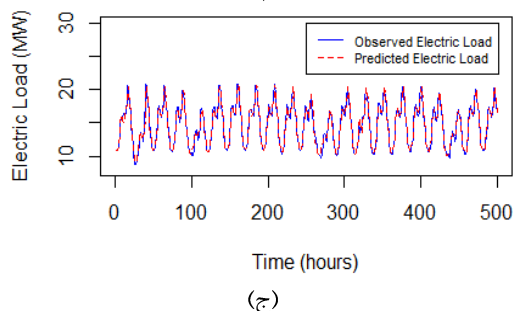
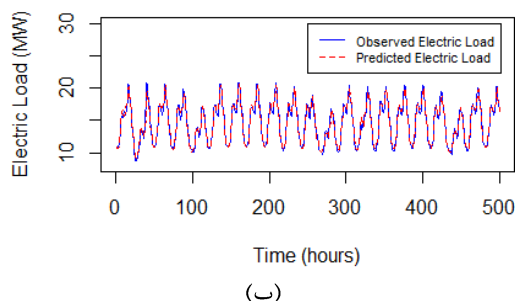
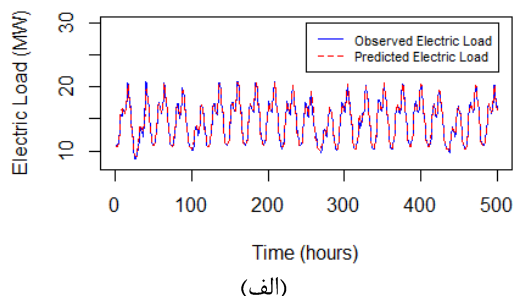
۳-۲- پیاده‌سازی روش‌ها

برای پیاده‌سازی روش‌ها از زبان برنامه‌نویسی متن‌باز R (نسخه 3.4.0) [۲۶] و بسته‌های Random Forest، e1071، SVM و MARS استفاده شده است [۲۷]. دقت روش‌های RF، SVM و MARS بر اساس معیارهای تعریف شده در بخش ۲-۵ محاسبه شده است. از روش اعتبارسنجی متقابل^{۱۵} برای بررسی عملکرد مدل‌ها بر اساس دسته‌بندی داده‌ها به صورت زیرمجموعه‌های آموزش و آزمون استفاده شده است. برای این منظور، ۸۵ درصد ابتدایی داده‌ها برای آموزش و از ۱۵ درصد باقیمانده برای آزمون در نظر گرفته شده است. برای هر سه روش پیاده‌سازی شده و در هر دو حالت پیش‌بینی کوتاه‌مدت و بلندمدت، شاخص‌های $RMSE$ ، MAE و R^2 برای زیرمجموعه‌های آموزش و آزمون جهت تنظیم بهینه پارامترها و همچنین ارزیابی نتایج استفاده شده است.

۳-۲-۱- آماده‌سازی داده‌های ورودی

در کلیه روش‌ها به منظور قابل مقایسه بودن نتایج و عملکرد روش‌ها از رویه یکسان برای آماده‌سازی داده‌های ورودی استفاده شده است. برای پیش‌بینی بار کوتاه‌مدت از داده‌های بار ساعت‌های قبلی به صورت متوالی به عنوان داده اصلی و از وضعیت هر روز (تعطیل یا غیر تعطیل بودن به صورت ۲۴ عدد ۰ یا ۱ برای کل ۲۴ ساعت)، دما و زاویه تابش خورشید در همان ساعت به عنوان داده‌های کمکی استفاده شده است. در حالت پیش‌بینی بلندمدت از داده‌های ساعت‌های مشابه در روزهای قبل به عنوان ورودی اصلی استفاده شده است. نحوه ورود داده‌های کمکی نیز مشابه حالت پیش‌بینی کوتاه‌مدت است. اما در مورد انتخاب تعداد داده‌های اصلی قبلی باید توجه داشت که طبعاً هر چه تعداد داده‌های قبلی بیشتر باشد، دقت پیش‌بینی بیشتر خواهد

پراکنش^{۱۷} مقادیر واقعی و پیش‌بینی شده برای این حالت ارائه شده است. با بررسی نتایج ارائه شده در این شکل‌ها مشخص است که RF نسبت به SVM و MARS نتایج بسیار بهتر و دقیق‌تری برای پیش‌بینی کوتاه‌مدت بار ارائه می‌دهد. این برتری در نمودارهای پراکنش شکل (۶) بسیار مشهودتر است. روش SVM نیز نسبت به MARS عملکرد بهتری دارد.



شکل (۵): مقایسه نتایج پیش‌بینی حالت کوتاه‌مدت؛ (الف) RF، (ب) SVM، (ج) MARS

۳-۴- نتایج پیش‌بینی بار بلندمدت

نتایج نمونه پیش‌بینی بار بلندمدت (مصرف ۲۴ ساعت بعد) سه روش در شکل (۷) نشان داده شده است. در شکل (۸) نیز نمودار پراکنش مقادیر واقعی و پیش‌بینی شده برای این حالت ارائه شده است. آنچه در این شکل‌ها مشهود است، کاهش دقت پیش‌بینی حالت بلندمدت نسبت به حالت کوتاه‌مدت است که طبعاً قابل پیش‌بینی است. به هر حال چالش پیش‌بینی در این حالت بسیار بیشتر است. با بررسی نتایج مشخص است که SMV نسبت به RF و MARS نتایج بهتری ارائه می‌دهد. این برتری در نمودارهای پراکنش شکل (۸) بیشتر مشهود است.

بود؛ اما باید توجه داشت که در صورت بالا بودن تعداد داده‌های قبلی، مسئله برازش بیش از حد^{۱۸} و نیز افزایش بار محاسباتی در فرآیند پیش‌بینی رخ خواهد داد. با توجه به تجربه به دست آمده در این تحقیق برای سه روش پیاده‌سازی شده، تعداد داده‌های قبلی برای هر دو حالت پیش‌بینی کوتاه‌مدت و بلندمدت به میزان ۴۸ عدد تنظیم شده که ضمن فراهم نمودن شرایط برای سرعت بالای آموزش، دقت پیش‌بینی مطلوبی را نیز ارائه می‌دهد.

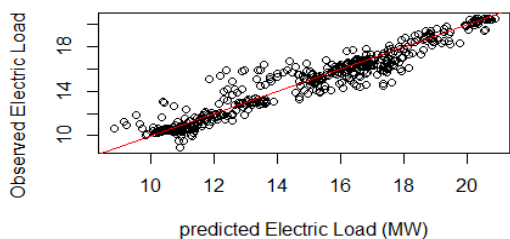
۳-۲-۲- تنظیمات پارامترهای مدل‌ها

برای ایجاد مدل RF باید سه پارامتر شامل تعداد حداقل گره‌های پایانه برای هر درخت ($nodesize$)، تعداد درخت‌ها در جنگل ($ntree$)، و تعداد متغیرهای انتخابی تصادفی برای رشد درخت ($mtry$) تنظیم شوند. پارامتر $nodesize$ اندازه هر درخت در جنگل را کنترل می‌کند و مقدار زیاد آن باعث کاهش بار محاسباتی و کاهش دقت خواهد شد. عکس این موضوع نیز در مورد مقادیر کم این پارامتر مصداق خواهد داشت. پارامتر $ntree$ تعداد درخت‌های تولید شده در یک مدل RF را تعیین می‌کند. مقدار بالای $ntree$ باعث افزایش دقت پیش‌بینی خواهد شد؛ اما از طرف دیگر منجر به افزایش زمان محاسبات نیز می‌گردد. پارامتر $mtry$ با ایجاد حالت تصادفی در فرآیند ساخت درخت‌ها بر دقت پیش‌بینی اثرگذار است. با افزایش این پارامتر، تعداد متغیرهای تصادفی در تشکیل درخت و نیز تشکیل جنگل افزایش یافته و در نتیجه دقت پیش‌بینی بالاتر خواهد رفت. انتخاب بالای این مقدار نیز منجر به افزایش بار محاسباتی خواهد شد. مقادیر به دست آمده برای پارامترهای $nodesize$ و $mtry$ بر اساس روش سعی و خطا و با هدف دستیابی به پاسخ‌های مطلوب در مرحله آموزش، به ترتیب برابر با ۵، ۳۰۰ و ۹ به دست آمده است.

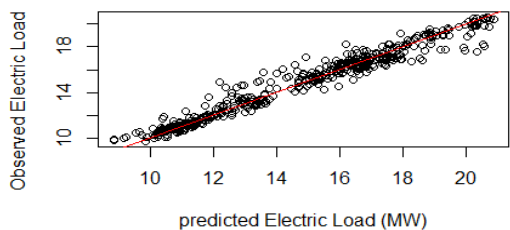
در روش SVM نیز سه پارامتر C ، γ و ϵ باید به خوبی تنظیم شوند تا نتایج مطلوبی به دست آید. راهکار مرسوم تعیین این پارامترها استفاده از اعتبارسنجی متقابل با هدف حداقل‌سازی خطای پیش‌بینی است. این پارامترها در این مقاله بر اساس تجارب کارهای قبلی و فرآیند سعی و خطا تنظیم شده‌اند. مقادیر تنظیمی به دست آمده برای این سه پارامتر به ترتیب برابر با ۱، ۰/۲۱ و ۰/۰۹۸ به دست آمده است. در ایجاد مدل MARS نیز سه پارامتر تنظیمی شامل حداکثر تعداد توابع پایه، پارامتر جریمه و حداکثر تراکنش بین متغیرها باید تعیین شوند. در این مقاله این سه پارامتر تنظیمی بر اساس تکرارهای مختلف با هدف حصول دقت مناسب، به ترتیب برابر با ۰/۷، ۴۴ و ۵ به دست آمده است. نتایج به دست آمده از پیاده‌سازی هر سه روش در هر دو حالت پیش‌بینی کوتاه‌مدت و بلندمدت بار در ادامه ارائه شده است.

۳-۳- نتایج پیش‌بینی بار کوتاه‌مدت

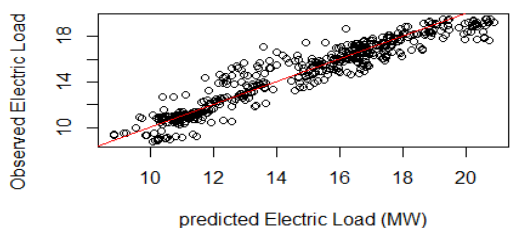
نتایج نمونه پیش‌بینی بار کوتاه‌مدت (مصرف ۱ ساعت بعد) برای سه روش در شکل (۵) نشان داده شده است. در شکل (۶) نیز نمودار



(الف)



(ب)



(ج)

شکل (۸): نمودار پراکنش در حالت پیش‌بینی بلندمدت: (الف) RF.

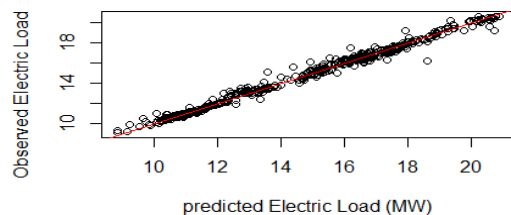
(ب) SVM، (ج) MARS (مقادیر بار بر حسب MW)

۳-۵- مقایسه کمی عملکرد روش‌ها

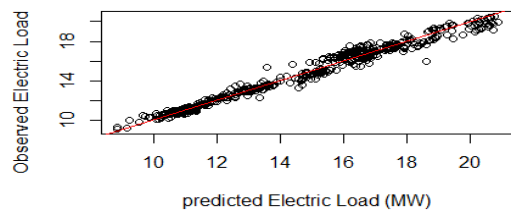
مقادیر عددی معیارهای ارزیابی خطای پیش‌بینی در جدول (۱) ارائه شده است. به منظور مقایسه بهتر نتایج، مقدار اختلاف نتیجه هر روش و هر معیار با بهترین عملکرد نیز بر حسب درصد محاسبه و درج شده است. ارزیابی و مقایسه عملکرد و دقت روش‌ها در حالت‌های مختلف با بررسی این جدول به خوبی قابل انجام است. نتایج به دست آمده نشان می‌دهد که در پیش‌بینی بار کوتاه‌مدت روش RF نتایج بسیار خوب و مطلوبی ارائه می‌دهد و در هر سه معیار با اختلاف قابل توجه نسبت به دو روش دیگر بهترین عملکرد را داشته است. روش SVM و MARS نیز به ترتیب رتبه بعدی را دارند و نکته جالب توجه این است که در مقایسه با RF اختلاف دقت مشابهی دارند.

اما در پیش‌بینی بار بلندمدت، روش SVM در تمامی معیارها با اختلاف قابل توجه، بهترین عملکرد را داشته است. نکته جالب توجه این است که در این حالت هر چند RF دقت کمتری نسبت به SVM ارائه می‌دهد اما اختلاف عملکرد آن در مقایسه با MARS به مقدار قابل توجهی بهتر است و این موضوع نشان‌دهنده عملکرد قابل قبول‌تر RF در مقایسه با MARS در پیش‌بینی بار بلندمدت است.

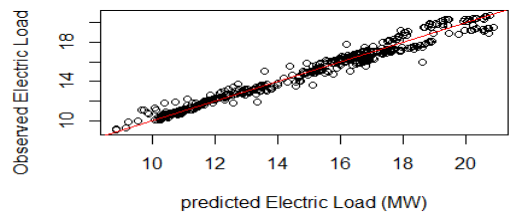
به صورت کلی می‌توان نتیجه گرفت که روش RF در پیش‌بینی بار کوتاه‌مدت عملکرد بسیار مطلوب‌تری ارائه می‌دهد. از طرفی روش SVM در پیش‌بینی بار بلندمدت دقت بسیار بهتری نسبت به دو روش



(الف)



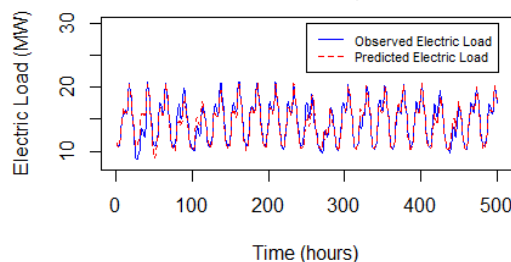
(ب)



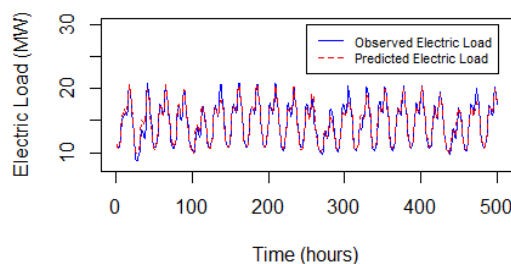
(ج)

شکل (۶): نمودار پراکنش در حالت پیش‌بینی کوتاه‌مدت: (الف) RF.

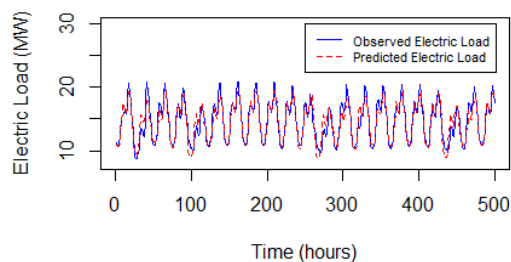
(ب) SVM، (ج) MARS (مقادیر بار بر حسب MW)



(الف)



(ب)



(ج)

شکل (۷): مقایسه نتایج پیش‌بینی حالت بلندمدت: (الف) RF، (ب)

SVM، (ج) MARS

جدول (۱): مقایسه عملکرد روش‌های پیش‌بینی پیاده‌سازی شده

| معیار | افق پیش‌بینی | RF | | SVM | | MARS | |
|----------------|--------------|--------|--------------------------|--------|--------------------------|--------|--------------------------|
| | | مقدار | درصد تفاوت نسبی با معیار | مقدار | درصد تفاوت نسبی با معیار | مقدار | درصد تفاوت نسبی با معیار |
| RMSE | کوتاه‌مدت | ۰/۳۲۱۴ | ۰/۰۰% | ۰/۴۶۹۸ | ۴۶/۱۷% | ۰/۵۱۱۲ | ۵۷/۰۵% |
| | بلندمدت | ۰/۸۶۲۳ | ۳۹/۲۶% | ۰/۶۱۹۲ | ۰/۰۰% | ۰/۹۴۸۵ | ۵۳/۱۸% |
| MAE | کوتاه‌مدت | ۰/۲۱۲۵ | ۰/۰۰% | ۰/۳۴۵۶ | ۶۲/۶۴% | ۰/۳۶۵۵ | ۷۲/۰۰% |
| | بلندمدت | ۰/۵۹۴۷ | ۴۰/۹۹% | ۰/۴۲۱۸ | ۰/۰۰% | ۰/۶۸۸۹ | ۶۳/۳۲% |
| R ² | کوتاه‌مدت | ۰/۹۹۵۶ | ۰/۰۰% | ۰/۹۸۸۱ | -۰/۷۵% | ۰/۹۸۶۴ | -۰/۹۲% |
| | بلندمدت | ۰/۹۶۱۳ | -۲/۳۶% | ۰/۹۸۴۵ | ۰/۰۰% | ۰/۹۵۲۶ | -۳/۲۴% |

دیگر دارد. و لذا روش RF و SVM به ترتیب برای پیش‌بینی بار الکتریکی کوتاه‌مدت و بلندمدت قابلیت بالایی از خود نشان می‌دهند.

۴- نتیجه‌گیری

در این مقاله سه روش مختلف یادگیری ماشین شامل RF، SVM و MARS برای هر دو حالت پیش‌بینی مصرف بار الکتریکی کوتاه‌مدت و بلندمدت پیاده‌سازی گردید. داده مورد استفاده، مصرف بار الکتریکی ساعتی بخشی از شهر همدان به همراه دمای هوا و زاویه ارتفاعی خورشید به عنوان ورودی‌های کمکی بوده است. دقت سه روش با سه معیار خطای مختلف سنجیده شد و مشخص گردید که روش جنگل تصادفی (RF) در پیش‌بینی بار کوتاه‌مدت و روش ماشین بردار پشتیبان (SVM) در پیش‌بینی بار بلندمدت بهترین عملکرد را با بیشترین دقت ارائه می‌کنند. در این بین روش اسپایلین‌های رگرسیون

تطبیقی چندمتغیره (MARS) در مقایسه با دو روش دیگر در هر دو افق پیش‌بینی و در تمامی معیارها کمترین دقت را دارا بود. و لذا با توجه به نتایج به دست آمده در این تحقیق، استفاده از روش RF برای پیش‌بینی بار کوتاه‌مدت و روش SVM برای پیش‌بینی بار بلندمدت توصیه می‌شود.

مراجع

- [۴] قدرت "نشریه مهندسی برق و الکترونیک ایران، سال چهاردهم، شماره سوم، صص. ۱۱۷-۱۰۹، پاییز ۱۳۹۶.
- [5] Song K. B., Baek Y. S., Hong D. H., and Jang G., "Short-term load forecasting for the holidays using fuzzy linear regression method", IEEE Transactions on Power Systems, Vol. 20, No. 1, pp. 96-101, Feb. 2005.
- [6] Selvarajoo S., Schlöpfer M. and Tan R., "Urban Electric Load Forecasting with Mobile Phone Location Data", Asian Conference on Energy, Power and Transportation Electrification (ACEPT), pp. 1-5, Nov. 2018.
- [7] Sudheer G. and Suseelatha A., "Short term load forecasting using wavelet transform combined with Holt-Winters and weighted nearest neighbor models", International Journal of Electrical Power & Energy Systems, Vol. 64, pp. 340-346, 2015.
- [8] Huang S. J. and Shih K. R., "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations", IEEE Transactions on Power Systems, Vol. 18, No. 2, pp. 673-679, May 2003.
- [9] Luo S., et. al., "Short-term Prediction of Electricity Consumption in Distribution Network during Peak Summer Period Based on ARIMA Model", IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), pp. 328-332, 2021.
- [10] Sharma S., Majumdar A., Elvira V., and Chouzenoux É., "Blind Kalman Filtering for Short-Term Load Forecasting", IEEE Transactions on Power Systems, Vol. 35, No. 6, pp. 4916-4919, Nov. 2020.
- [11] Hoori A. O., Kazzaz A. A., Khimani R., Motai Y., and Aved A. J., "Electric Load Forecasting Model Using a Multicolumn Deep Neural Networks", IEEE Transactions on Industrial Electronics, Vol. 67, No. 8, pp. 6473-6482, Aug. 2020.
- [12] Fahiman F., Erfani S. M., Rajasegarar S., Palaniswami M., and Leckie C., "Improving load forecasting based on deep learning and K-shape clustering", International Joint Conference on Neural Networks (IJCNN), pp. 4134-4141, 2017.
- [13] Kong Z., Zhang C., Lv H., Xiong F., and Fu Z., "Multimodal Feature Extraction and Fusion Deep Neural Networks for Short-Term Load Forecasting", IEEE Access, Vol. 8, pp. 185373-185383, 2020.
- [14] Chen K., Chen K., Wang Q., He Z., Hu J., and He J., "Short-Term Load Forecasting With Deep Residual Networks", IEEE Transactions on Smart Grid, Vol. 10, No. 4, pp. 3943-3952, Jul. 2019.
- [15] Ouyang T., He Y., Li H., Sun Z., and Baek S., "Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network", IEEE Transactions on
- [1] Nti, I. K., Teimeh M., Nyarko-Boateng O., and Adebayo F. A., "Electricity load forecasting: a systematic review", Journal of Electrical Systems and Information Technology, Vol. 7, No. 1 pp: 1-19, 2020.
- [۲] کی نیا فرشید، صیادی شهرکی فهیمه، اسماعیلی سعید، کنترل ولتاژ و توان راکتیو دو مرحله‌ای در شبکه‌های توزیع در حضور سیستم‌های فتوولتائیک مبتنی بر پیش‌بینی کوتاه مدت بار و تولید، نشریه مهندسی برق و الکترونیک ایران، سال پانزدهم، شماره اول، صص. ۴۳-۵۴، بهار ۱۳۹۷.
- [۳] کریمی مازیار، کرمی حسین، غلامی مصطفی، خطیب‌زاده هادی، مسلمی نیکی، "اولویت‌بندی روزهای مشابه جهت پیش‌بینی بار کوتاه‌مدت شبکه ایران با در نظرگیری دما و بخش‌بندی سیستم

¹⁰ Generalized Gaussian Radial Basis Function¹¹ Solar Altitude Angle¹² Root Mean Square Error¹³ Mean Absolute Error¹⁴ Coefficient of Determination¹⁵ Cross-Validation Technique¹⁶ Over Fitting¹⁷ Scatter Plot

- Emerging Topics in Computational Intelligence, Vol. 3, No. 2, pp. 127-136, Apr. 2019.
- [16] Ali M., Adnan M., Tariq M., and Poor H. V., "Load Forecasting Through Estimated Parametrized Based Fuzzy Inference System in Smart Grids", IEEE Transactions on Fuzzy Systems, Vol. 29, No. 1, pp. 156-165, Jan. 2021.
- [17] Liu F., Dong T., Hou T., and Liu Y., "A Hybrid Short-Term Load Forecasting Model Based on Improved Fuzzy C-Means Clustering, Random Forest and Deep Neural Networks", IEEE Access, Vol. 9, pp. 59754-59765, 2021.
- [18] Kandil M. S., El-Debeiky S. M., and Hasanien N. E., "Long-term load forecasting for fast developing utility using a knowledge-based expert system", IEEE Transactions on Power Systems, Vol. 17, No. 2, pp. 491-496, May 2002.
- [19] Mosavi, A., Salimi M., Faizollahzadeh Ardabili S., Rabczuk T., Shamshirband S., and Varkonyi-Koczy A. R., "State of the art of machine learning models in energy systems, a systematic review", Energies, Vol. 12, No. 7, 2019.
- [20] Breiman, L., "Random forests", Machine learning, Vol. 45, No. 1 pp. 5-32, 2001.
- [21] Grömping U., "Variable importance assessment in regression: linear regression versus random forest", The American Statistician, Vol. 63, No. 4, pp. 308-319, 2009.
- [22] Wang, Z., Wang Y., Zeng R., Srinivasan R. S., and Ahrentzen S., "Random Forest based hourly building energy prediction", Energy and Buildings Vol. 171 pp. 11-25, Jul. 2018.
- [23] Kisi, O., and Cimen M., "Precipitation forecasting by using wavelet-support vector machine conjunction model", Engineering Applications of Artificial Intelligence Vol. 25, No. 4, pp. 783-792, 2012.
- [24] Yoon, H., Jun S. C., Hyun Y., Bae G. O., and Lee K. K., "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer", Journal of hydrology, Vol. 396, No. 1-2, pp. 128-138, 2011.
- [25] Zhang, H., and Singer B. H., Recursive partitioning and applications, Springer Science & Business Media, 2010.
- [26] Al-Musaylh, M. S., Deo R. C., Adamowski J. F., and Li Y., "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia", Advanced Engineering Informatics, Vol. 35 pp. 1-16, Jan. 2018.
- [27] Kuen, T. H., Ramsey J. W., and Threlkeld J. L., Thermal Environmental Engineering, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [28] <https://www.r-project.org/>, Last visit: August 23, 2021.
- [29] Lantz, B., Machine learning with R: expert techniques for predictive modeling, Packt publishing Ltd., 2019.

زیر نویس‌ها

¹ Exponential Smoothing² Autoregressive Moving Average³ Autoregressive Integrated Moving Average⁴ Random Forest⁵ Support Vector Machine⁶ Multivariate Adaptive Regression Spline⁷ Classification and Regression Tree⁸ Boot Strap⁹ Out-of-bag