

مدلسازی داده‌رانه مدت زمان تداوم وقفه در شبکه توزیع برق با در نظر گرفتن نگهداری و تعمیرات پیشگیرانه، و تحلیل اثر آن

نیما سالک گیلانی^۱ علیرضا فریدونیان^۲

۱- دانشجوی دکتری- دانشکده مهندسی برق- دانشگاه صنعتی خواجه نصیرالدین طوسی- تهران- ایران

nima.salek@email.kntu.ac.ir

۲- استادیار- دانشکده مهندسی برق- دانشگاه صنعتی خواجه نصیرالدین طوسی - تهران- ایران

fereidunian@eetd.kntu.ac.ir

چکیده: یکی از معیارهای مهم در سنجش عملکرد شرکت‌های توزیع برق، میزان رضایت مشتریان است. از دید اکثر مشتریان مهمترین معیار عملکرد این شرکت‌ها، مدت زمان تداوم وقفه‌ها است. یکی از ابزاری‌های شرکت‌های توزیع برق برای کنترل و کاهش مدت زمان تداوم وقفه‌ها نگهداری و تعمیرات پیشگیرانه است. این پژوهش به مدلسازی مدت زمان تداوم وقفه با در نظر گرفتن نگهداری و تعمیرات پیشگیرانه می‌پردازد. ابتدا داده‌های وقفه‌های سیستم از داشبورد بهره‌برداری شرکت توزیع نیروی برق تهران بزرگ استخراج گشت و سازماندهی شد و سپس با داده‌های هواشناسی در بازه زمانی مورد بررسی تلفیق گردید. در ادامه پس از پاکسازی داده‌ها از روش‌های رگرسیون خطی و غیرخطی برای مدلسازی مدت زمان تداوم وقفه استفاده شده است. نوآوری‌ها و مشارکت‌های این پژوهش را می‌توان: (۱) مدلسازی داده‌رانه‌ی زمان تداوم وقفه با استفاده از الگوریتم‌های مختلف خطی و غیرخطی، (۲) در نظر گرفتن اثر محل استقرار تیم تعمیرات و ترافیک شهری به صورت مدت زمان رسیدن تیم تعمیرات به محل حادثه، (۳) در نظر گرفتن تاریخچه انجام عملیات نگهداری پیشگیرانه در مدلسازی زمان تداوم وقفه و تحلیل حساسیت زمان تداوم وقفه نسبت به عملیات نگهداری پیشگیرانه، و (۴) تحلیل تاثیر ویژگی‌های ورودی بر پیش‌بینی زمان تداوم وقفه دانست.

واژه‌های کلیدی: تعمیرات و نگهداری پیشگیرانه، مدت زمان تداوم وقفه، شبکه توزیع برق، داده‌کاوی، مدلسازی

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.19.1.1

تاریخ ارسال مقاله: ۱۴۰۰/۰۴/۱۳

تاریخ پذیرش مشروط مقاله: ۱۴۰۰/۰۵/۱۰

تاریخ پذیرش مقاله: ۱۴۰۰/۰۶/۱۶

نام نویسنده‌ی مسئول: دکتر علیرضا فریدونیان

نشانی نویسنده‌ی مسئول: تهران - خیابان شریعتی- ضلع جنوب شرقی پل سید خندان - دانشگاه صنعتی خواجه نصیرالدین طوسی - دانشکده

مهندسی برق - کدپستی ۱۶۳۱۷۱۴۱۹۱

۱- مقدمه

وقوع وقفه در شبکه‌های توزیع امری متداول است، که در برخی کشورها به علت پیری زیرساخت‌ها و تغییرات آب و هوا رو به افزایش است [۱-۳]. این وقفه‌های بی‌برنامه می‌تواند منجر به قطع سرویس طولانی مدت و نارضایتی قابل توجه مشترکین شود. بنابراین کاهش تعداد و زمان این وقفه‌ها برای اکثر شرکت‌های خدماتی یک اولویت است. با این حال، وقفه‌های بی‌برنامه نامنظم بوده و پیش‌بینی آن‌ها دشوار است [۴]. بنابراین کاهش تعداد وقفه‌های بی‌برنامه و مدیریت مدت زمان تداوم آن در اکثر شرکت‌های توزیع در اولویت قرار دارد [۵]. حرکت در این جهت نیازمند شناخت دقیق عملکرد سیستم و استفاده از تجارب گذشته است. برای کسب تجربه به سوابق و داده‌های عملکرد پیشین سیستم نیاز است. استخراج روایت اعداد از اتفاقات گذشته جهت کمک برای تفسیر آن‌ها به کمک ابزارهای داده‌کاوی امکان‌پذیر است. ابزارهای داده‌کاوی به پژوهشگران کمک می‌کند تا از سرمایه اطلاعاتی داده‌های موجود بهره‌مند شوند و به پیش‌بینی و مدلسازی اتفاقات آینده بپردازند.

در زمینه پیش‌بینی و مدلسازی تعداد وقفه‌های بی‌برنامه و همینطور بررسی عوامل موثر بر آن، بر خلاف مدت زمان تداوم وقفه، تحقیقات گسترده‌ای صورت پذیرفته است. یکی از عوامل موثر در تعداد و زمان تداوم وقفه‌های بی‌برنامه، نحوه نگهداری و تعمیرات (نت) در شبکه توزیع است. با بلوغ سیستم‌های نت و همچنین گسترش سطح تولید و شبکه در طی چند دهه اخیر، سازمان‌ها و شرکت‌های زیادی در جهان به دنبال بررسی طول عمر باقی‌مانده تجهیزات [۶-۷] و به‌کارگیری سیستم‌های نت و تغییر روند گذشته خویش برآمدند [۸-۹]. انگیزه این پژوهش مدلسازی مدت زمان تداوم وقفه با در نظر گرفتن اثر نت پیشگیرانه بر آن است.

پژوهش در رابطه با مدت زمان تداوم وقفه زمینه‌های مختلفی را شامل می‌شود. این زمینه‌ها شامل مدلسازی آماری مدت زمان تداوم وقفه و پیش‌بینی مدت زمان تداوم وقفه می‌شود.

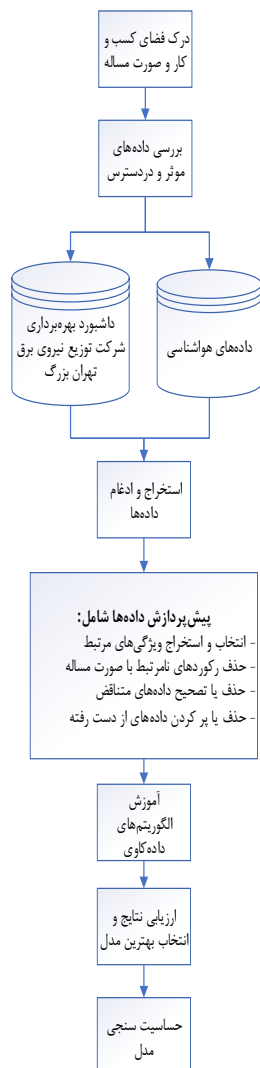
در زمینه مدلسازی آماری مدت زمان تداوم وقفه، هدف یافتن مناسب‌ترین تابع توزیع چگالی احتمال با استفاده از داده‌های گذشته، جهت مدلسازی رفتار بلند مدت آتی شبکه است. در این زمینه در [۱۰] که یک مدل وایبول-مارکوف تصادفی برای تحلیل پایایی سیستم توزیع پیشنهاد شده است، از فرض یک توزیع نمایی برای مدلسازی تداوم وقفه استفاده شده است. در [۱۱] با استفاده از توزیع وایبول نشان داده شده است که بی‌توجهی به جنبه‌های احتمالاتی تداوم وقفه می‌تواند منجر به خطا تا ۴۰ درصد در پیش‌بینی هزینه‌های وقفه مورد انتظار گردد. در [۱۲] به بررسی توانایی توزیع بتا برای مدل‌سازی شاخص‌های پایایی می‌پردازد. در این پژوهش با اشاره به مزیت توصیف یک شاخص با توزیع احتمالاتی آن به جای استفاده از یک عدد تنها، که معمولاً میانگین آن شاخص است، به مقایسه توزیع بتا با ۷ توزیع

دیگر پرداخته شده است. این مطالعه در مجموع این توزیع احتمالاتی را توزیعی انعطاف‌پذیر و مناسب برای مدلسازی تداوم وقفه و شاخص‌های پایایی معرفی می‌کند. در [۱۳] به مدلسازی تداوم وقفه خطوط هوای اولیه، با استفاده از داده‌های ۸ ساله از یک شرکت توزیع در شمال شرقی آمریکا پرداخته شده است. در این پژوهش تنها مدل‌های بور و وایبول مورد بررسی قرار گرفته‌اند و پارامترهایشان استخراج شده است. در مرجع [۱۴] هدف مدلسازی احتمالاتی مدت زمان تداوم وقفه در شبکه توزیع، و استخراج تابع نگهداری‌پذیری با استفاده از مدل‌های احتمالاتی موجود و به کمک داده‌های خاموشی شرکت توزیع نیروی برق تهران بزرگ است. بدین منظور داده‌های خاموشی‌های بی‌برنامه در سطوح ولتاژ فشار ضعیف و فشار متوسط استخراج گشتند. سپس این داده‌ها با ۱۱ توزیع احتمالاتی برازش شدند و در انتها نتایج هر یک از با استفاده از معیارهایی ارزیابی و گزارش شد. با اینکه در ادبیات پیشین برای مدلسازی این شاخص‌ها توصیه به استفاده از توزیع‌های بور و وایبول شده است، نتایج برازش ۱۱ توزیع احتمالاتی در این پژوهش نشان می‌دهد، توزیع مقدار حدی تعمیم یافته^۱ بهترین برازش و عملکرد را دارا است. توزیع مقدار حدی تعمیم یافته در هر سه سال مورد مطالعه از دید شاخص‌های ارزیابی عملکرد جایگاه نخست را کسب کرده و به عنوان توزیع پیشنهادی این مطالعه ارائه شده است.

در زمینه پیش‌بینی مدت زمان تداوم وقفه، بر خلاف مبحث پیش‌بینی تعداد وقفه، مطالعات کمی صورت پذیرفته است. همانطور که می‌توان پژوهش‌ها را در زمینه پیش‌بینی تعداد وقفه به دو زیر گروه شرایط آب و هوایی سخت [۱۵-۱۸] و شرایط نرمال [۱۹-۲۲] تقسیم کرد، در زمینه پیش‌بینی مدت زمان تداوم وقفه نیز دسته‌ای از پژوهش‌ها به بررسی این موضوع در شرایط طوفانی پرداختند [۲۳-۲۴]. در زمینه پیش‌بینی مدت زمان تداوم وقفه در شرایط عادی، در مرجع [۲۵] پژوهشگران اهمیت آماری تعدادی از ویژگی‌ها بر روی مدت زمان تداوم وقفه بررسی کردند، اما الگوریتم پیش‌بینی خاصی ارائه ندادند. مرجع [۲۶] با فرض خاموشی نیروگاه و تجزیه فرآیند بازگشت شبکه به فعالیت‌های منفرد سعی در پیش‌بینی زمان وقفه دارد. در این پژوهش فرآیندهای بازگشت نیروگاه در کنار فرآیندهای کلیدزنی برای بازگشت شبکه دیده شده است. برای هر فرآیند سه زمان محتمل، خوشبینانه و بدبینانه در نظر گرفته شده است و به کمک روش مسیر بحرانی تقدم و قابلیت انجام همزمان فرآیندها در نظر گرفته شده است. در مرجع [۲۷] تکنیکی بر پایه منطق فازی طراحی شده است که برای تعیین اهمیت نسبی ویژگی‌ها به دانش فرد خبره متکی است. روش پیشنهادی پیش‌بینی آماری را با دانش مهندسی ترکیب می‌کند. کالیبره کردن این ترکیب دشوار است و همچنین خروجی روش برای گزارش به مشترکین مناسب نمی‌باشد و برای استفاده در بهینه‌سازی دیگری توسعه داده شده است. مرجع [۲۸] به پیش‌بینی مدت زمان تداوم وقفه به کمک گزارشات خاموشی

نباشد. از طرفی باید به دنبال پست فوق توزیعی گشت که کمترین تغییر در تعداد فیدر را در بازه چهار ساله مورد مطالعه داشته باشد. پس از جستجو در بین داده‌ها یک پست فوق توزیع در این منطقه که در بازه مورد مطالعه تغییری در تعداد فیدرهای آن اتفاق نیافتاده است انتخاب گشت. در مرحله بعد در پایگاه داده باید به دنبال ویژگی‌هایی بود، که در پیش‌بینی زمان وقفه کمک کننده باشند.

از میان ویژگی‌های قابل دسترس در داشبورد بهره‌برداری ۱۶ ویژگی استخراج گشتند. این ۱۶ ویژگی شامل متوسط درصد بارگذاری پست‌ها، KVA کل ترانس‌های فیدر، متوسط درصد نامتعادلی پست‌ها، سن متوسط ترانس‌های فیدر، ساعت وقفه، مدت زمان رسیدن تیم نت، TLM^۲، علت قطع، علت اشکال، گروه قطع، انرژی تامین نشده، تعداد پست فیدر، تعداد ترانس فیدر، جریان قطع اولیه (مقدار جریانی که در اثر وقفه روی داده قطع شده است)، شماره روز در سال و نام فیدر فشار متوسط می‌شود. این ویژگی‌ها به همراه ۸ ویژگی آب و هوایی به عنوان ورودی‌های اولیه جهت تحلیل و پیش‌پردازش انتخاب و استخراج



شکل (۱): فرآیند پژوهش

و تعمیرات و با استفاده از روش‌های یادگیری ماشینی و پردازش زبان طبیعی می‌پردازد. در این پژوهش از عوامل محیطی بیرونی (مانند سرعت باد و شرایط آب و هوایی دیگر، ساعت روز)، ویژگی‌های فیزیکی و دانش مهندسی که در تاریخچه گزارش‌های خاموشی و تعمیرات موجود است، جهت پیش‌بینی زمان وقفه استفاده شده است. در ابتدا پیش‌بینی اولیه‌ای از ویژگی هدف با اطلاعات اولیه و به کمک شبکه عصبی پیش‌خور انجام می‌شود. در ادامه پس از رسیدن هر گزارش تعمیر، متن گزارش با کمک یک شبکه عصبی بازگشتی به صورت نمایش فضای برداری خلاصه می‌شود تا با یکپارچه شدن با ویژگی‌های فیزیکی به پیش‌بینی زمان وقفه کمک کنند. شبکه عصبی بازگشتی دیگری وظیفه برورسانی تدریجی زمان تداوم وقفه پیش‌بینی شده را با استفاده از اطلاعات جدید به عهده دارد. در مرجع [۲۹] با اطلاعات خرابی در سه ماهه اول سال ۱۳۹۷ حوزه معاونت شمال شرق تهران بزرگ که شامل بیش از هفت هزار رکورد ثبت شده است به پیش‌بینی مدت زمان تداوم وقفه پرداخته شده است. از یک سو بازه زمانی کوتاه اجازه بررسی چرخه حوادث را نمی‌دهد و از سوی دیگر وسعت ناحیه مورد بررسی پارامترهای زیادی را درگیر خواهد کرد که فرآیند مدلسازی را سخت خواهد می‌کند.

کارهای پژوهشی زیادی در زمینه پیش‌بینی مدت زمان وقفه انجام نشده است. همچنین در مطالعات انجام شده جای خالی بررسی اثر زمان انجام نت پیشگیرانه، و نیز اثر محل حضور تیم نت، و ترافیک شهری مشهود است. در این پژوهش سعی بر آن بوده است تا با در نظر گرفتن پارامترهای ذکر شده به مدلسازی مدت زمان تداوم وقفه و بررسی اهمیت پارامترهای تاثیرگذار پرداخته شود. نوآوری‌ها و مشارکت‌های این مقاله را می‌توان به شرح زیر خلاصه کرد:

- ۱- ارایه یک مدل داده‌رانه‌ی زمان تداوم وقفه با استفاده از الگوریتم‌های مختلف خطی و غیرخطی
- ۲- در نظر گرفتن اثر محل استقرار تیم تعمیرات و ترافیک شهری به صورت مدت زمان رسیدن تیم تعمیرات به محل حادثه در مدل زمان تداوم وقفه
- ۳- در نظر گرفتن زمان انجام عملیات نگهداری پیشگیرانه در مدل زمان تداوم وقفه و تحلیل حساسیت زمان تداوم وقفه نسبت به عملیات نگهداری پیشگیرانه
- ۴- تحلیل اهمیت شاخص‌ها بر اساس اثرگذاری بر زمان تداوم وقفه

۲- روش پیشنهادی

در این پژوهش برای جمع‌آوری داده‌ها از شرکت توزیع نیروی برق تهران بزرگ سعی بر آن بوده است که داده‌ها تا حد امکان محدود شوند. این محدودیت بدان جهت است که این پژوهش درگیر ویژگی‌های غیر قابل استخراجی مانند تفاوت در مهارت و سرعت اکیپ‌های نت، شرایط خاص منطقه‌ای و رویکرد متفاوت مدیریتی

شدند. اطلاعات آب و هوایی شامل دما، سرعت باد، نقطه شبنم، دید افقی، فشار هوا، میزان بارش در ۶ ساعت گذشته، بیشینه و کمینه دمای آن روز می‌باشد. فرایند پژوهش در شکل ۱ به تصویر کشیده شده است.

۱-۲- پیش پردازش داده‌ها

برای استخراج داده‌ها فرایند محتاطانه و زمان‌بری پیاده‌سازی شد. این فرایند شامل فیلتر کردن ستون‌های گروه قطع و علت قطع در خاموشی‌های بی‌برنامه، محدود کردن خاموشی‌های بی‌برنامه به نت پیشگیرانه، محاسبه ویژگی‌هایی مانند سن متوسط ترانس‌های هر فیدر و زمان رسیدن تیم نت به محل با اطلاعات موجود، پاکسازی داده‌ها و جمع‌آوری داده‌های آب و هوایی با داده‌های داشبورد بهره‌برداری می‌شود. در ویژگی گروه قطع رکوردهایی که شامل موارد غیر مرتبط با مساله نت می‌باشند، مانند روشنایی معابر و انشعابات مشترکین، از پایگاه داده مورد بررسی حذف شده‌اند. در ستون علت قطع نیز برای محدودسازی وقفه‌ها در خاموشی‌های بی‌برنامه هر فیدر، به مواردی که با نت و بهره‌برداری شبکه در ارتباط هستند، باید بررسی و فیلتر شود. عواملی مانند سرقت شبکه، سرقت تجهیزات، کلنگ خوردگی، برخورد اشیاء خارجی، آتش سوزی، عبور حیوانات، عیب نامشخص و غیره که دلایلی جدای از ساختار و نحوه بهره‌برداری شبکه توزیع دارند از داده‌ها حذف گشت. در این پژوهش پست‌های توزیع به نوع خاصی محدود نشده است، اما اکثر پست‌ها از نوع زمینی هستند. مقدار ویژگی TLM به معنای فاصله زمانی (تعداد روز) میان آن وقفه تا آخرین نت پیشگیرانه انجام شده برای آن فیدر است و تاثیر زمان انجام عملیات نت پیشگیرانه را مدل می‌کند.

علت اشکال به عنوان یک ویژگی که بصورت جزئی‌تر به توصیف علت قطع یا به عبارتی مشکل به وجود آمده می‌پردازد، دارای تعداد بالایی داده از دست رفته^۳ (تقریباً ۳۴ درصد) بود. به علت اهمیت اطلاعات این ویژگی، ستون آن حذف نشده و مقادیر از دست رفته با عبارت missing جایگزین گشت. رکوردهایی که دارای زمان تداوم وقفه زیر ۵ دقیقه (۵ رکورد) هستند، به دلیل جای‌گیری در دسته وقفه‌های گذرا و کم اهمیت بودن پیش‌بینی این دسته حذف گشتند. رکوردهایی با انرژی توزیع نشده برابر صفر (۴ عدد) بخاطر ثبت اشتباه یا عدم ارتباط داده با مساله مورد بررسی حذف شد.

در کنار پاکسازی داده‌ها و افزودن ویژگی‌های جدید، جمع‌آوری داده‌ها فرآیندیست که با توجه به فرمت ذخیره‌سازی و فرکانس نمونه‌برداری در هر یک از پایگاه‌های داده باید صورت پذیرد. داده‌های آب و هوایی در بازه‌های زمانی ۳ ساعته ذخیره شده بودند. برای هر وقفه اطلاعات نزدیکترین رکورد زمانی ثبت شده به عنوان اطلاعات آب و هوایی با ویژگی‌های موجود آن وقفه جمع‌گشت.

در ادامه فرایند پژوهش پس از پاکسازی و جمع‌آوری داده‌ها، از روش‌های رگرسیون خطی و غیرخطی برای این پیش‌بینی استفاده

شده است. این روش‌ها شامل رگرسیون خطی، رگرسیون بردار پشتیبان^۴ با کرنل‌های خطی، چند جمله‌ای، گوسی و همچنین جنگل تصادفی^۵ می‌شود.

۲-۲- روش‌های مدلسازی داده‌رانه

روش‌های داده‌رانه بر ساختن سیستمی تمرکز دارند که می‌تواند پاسخ صحیح را بر پایه مشاهده و آموزش با تعداد زیادی نمونه داده، تشخیص دهد. روش‌های مدل‌رانه، برخلاف آن، با استفاده از قوانین صریح تلاش بر بدست آوردن دانش و تصمیم‌گیری دارند. در ادامه هر یک از پنج روش استفاده شده برای مدلسازی داده‌رانه، شامل سه روش اصلی که یکی از آنان با سه کرنل مختلف است، به اختصار شرح داده شده‌اند.

۱-۲-۲- رگرسیون خطی

در رگرسیون خطی فرض بر این است که می‌توان متغیر خروجی را بر حسب ترکیب خطی متغیرهای ورودی بصورت تقریبی مدل کرد. فرض کنید مجموعه داده‌های آموزش بصورت $\{(x_1, z_1), \dots, (x_l, z_l)\}$ باشد، که در آن $x_i \in R^n$ بردار ویژگی‌های ورودی و $z_i \in R$ مقدار خروجی یا هدف است. با این فرض‌ها فرم استاندارد رگرسیون خطی بصورت زیر خواهد بود [۳۰]:

$$f_{w,b}(x) = wx + b \quad (1)$$

در معادله ۱ بردار w بردار ضرایب و n بعدی و b یک عدد حقیقی می‌باشد. برای محاسبه پارامترها در رگرسیون خطی از روش حداقل مربعات استفاده می‌شود. به دلیل سادگی این روش، امکان مدلسازی داده‌هایی که به صورت غیرخطی جدایی‌پذیر هستند، میسر نمی‌باشد. با توجه به فضای غیرخطی ورودی‌ها نمی‌توان از آن انتظار بالایی داشت.

۲-۲-۲- رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان براساس مفاهیم ماشین‌های بردار پشتیبان^۶ بنا شده است و تا کنون برای پیش‌بینی در زمینه‌های مختلف به کار برده شده است. با فرض‌هایی مشابه آنچه برای رگرسیون خطی در نظر گرفته شده است فرم استاندارد مساله رگرسیون بردار پشتیبان بصورت زیر خواهد بود [۳۱-۳۲]:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \\ \text{subject to} \quad & w^T \phi(x_i) + b - z_i \leq \varepsilon + \xi_i, \\ & z_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l. \end{aligned} \quad (2)$$

در معادلات ۲ پارامترهای C و ε به ترتیب با عناوین پارامتر ضریب جریمه یا تنظیم و پارامتر تنظیم ناحیه عدم حساسیت شناخته می‌شوند. در این معادلات نمونه‌هایی که پیش‌بینی آن‌ها بیش از مقدار

نیز صورت می‌پذیرد. هر درخت مقداری را برای نمونه جدید پیش‌بینی می‌کند و خروجی الگوریتم میانگین ساده یا وزن‌دار این مقادیر خواهد بود.

این الگوریتم عملکرد تخمینی بهتری نسبت به یک درخت تنها دارد. هر درخت تخمینگر به تنهایی بایاس پایین و واریانس بالا دارد در صورتیکه تجمیع آن‌ها منجر به مصالحه‌ای بین بایاس و واریانس خواهد شد. این الگوریتم برای مسائلی با ابعاد بالا به خوبی کار کرده و می‌توان از آن در مسائل نامتعادل^۱ نیز بهره برد. علاوه بر این، جنگل تصادفی معیاری از اهمیت متغیرها را با توجه به خروجی پیش‌بینی ارائه می‌دهد [۳۳].

۳-۲-۲- مدلسازی جهت پیش‌بینی مدت زمان تداوم وقفه

در ابتدا داده‌های موجود برای فرآیند آموزش و آزمایش بصورت تصادفی به نسبت هشتاد به بیست تفکیک گشتند. برای فرآیند آموزش از روش اعتبارسنجی متقابل کی-لايه‌ای^{۱۱} استفاده شده است. این روش یکی از روش‌های توسعه و ارزیابی قابلیت تعمیم یک الگوریتم پیش‌بین است. در این روش داده‌ها به طور تصادفی به k گروه تقریباً هم اندازه تقسیم می‌شوند. هر بار یک گروه به عنوان مجموعه ارزیابی و $k-1$ گروه دیگر به عنوان مجموعه آموزش در نظر گرفته می‌شوند. انتخاب مقدار k در این روش با مصالحه بین بایاس و واریانس در ارتباط است. به طور معمول، با توجه به این ملاحظات، مقدار k عدد ۵ یا ۱۰ در نظر گرفته می‌شود، زیرا به طور تجربی اثبات شده است که با انتخاب این مقادیر تخمین میزان خطای آزمون دچار بایاس یا واریانس بیش از حد نخواهد بود [۳۴]. در این پژوهش برای k مقدار ۵ در نظر گرفته شده است. برای مقیاس^{۱۲} کردن داده‌ها دو روش متداول استانداردسازی^{۱۳} و نرمال‌سازی^{۱۴} داده‌ها وجود دارد [۳۰]. در روش استاندارد سازی ابتدا میانگین داده‌ها از آن‌ها کم شده و سپس بر واریانشان تقسیم می‌شوند تا داده‌های جدید دارای واریانس واحد باشند. در روش نرمال‌سازی ابتدا مینیمم داده‌ها از آن‌ها کم شده و سپس بر طول بازه داده‌ها تقسیم می‌شوند تا داده‌های جدید در بازه ۰ و ۱ توزیع شوند. روش استانداردسازی در مقابل داده‌های دورافتاده مقاوم‌تر است، ازینرو در این پژوهش این روش بر روی داده‌ها اعمال شده است.

۳- نتایج مدل‌سازی

بعد از تلفیق، پاکسازی و پیش‌پردازش داده‌ها نوبت به اعمال الگوریتم‌های پیش‌بینی می‌رسد. برای این کار در ابتدا بیست درصد از نمونه‌ها به صورت تصادفی انتخاب شدند و به عنوان داده تست بکار گرفته شدند. به عبارتی از ۶۲۷ رکورد خروجی مرحله پیش‌پردازش ۱۲۶ نمونه به عنوان داده تست و ۵۰۱ نمونه به عنوان داده آموزش بصورت تصادفی انتخاب شدند. از آنجایی رسم تمام مقادیر خروجی

ε با مقدار حقیقی اختلاف باشند جریمه خواهند شد. میزان این جریمه بسته به اینکه مقدار واقعی برای نمونه در بالا یا پایین خط پیش‌بینی باشد، توسط ξ^+ و ξ^- تعیین می‌شود. برای حل این مساله بهینه‌سازی، دوگان آن بصورت زیر تعریف می‌شود:

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \\ & \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l z_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & e^T (\alpha - \alpha^*) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l. \end{aligned} \quad (۳)$$

که در آن برداری است که تمامی مقادیر آن یک می‌باشد، Q ماتریس l در l نیمه معین است که بصورت زیر تعریف می‌شود:

$$Q_{ij} = K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j) \quad (۴)$$

در معادله ۴، K تابع کرنل است. در کرنل خطی امکان جداسازی داده‌های همراه با نویز و همچنین داده‌هایی که به صورت غیرخطی جدایی پذیر هستند، میسر نمی‌باشد. در این شرایط معمولاً به جای استفاده از کلاسه‌بندی غیرخطی در فضای با ابعاد کوچک، داده‌ها را به یک فضا با ابعاد بیشتر نگاشت کرده و سپس آن‌ها را به کمک یک جداسازی خطی کلاسه‌بندی می‌کنند. هر تابع کرنل دارای یک یا چند پارامتر تنظیم می‌باشد. سه نمونه از توابع کرنل کاربرد شامل خطی، چند جمله‌ای و گوسی می‌شوند و هر یک پارامترهای خاص خود را دارند.

پس از حل معادله بهینه‌سازی تابع تقریب زده شده به فرم زیر خواهد بود:

$$\sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(x_i, x) + b \quad (۵)$$

یکی از چالش‌های استفاده از رگرسیون بردار پشتیبان تنظیم مقدار پارامترهای آن است. این پارامترها شامل مقادیر C و ε در کنار پارامترهای توابع کرنل است. برای تنظیم این پارامترها از جستجوی شبکه‌ای^{۱۵} بهره برده شده است.

۳-۲-۲- جنگل تصادفی

الگوریتم جنگل تصادفی در سال ۲۰۰۱ توسط بریمن^{۱۶} معرفی گشت [۳۳]. جنگل تصادفی، در واقع گروهی از درخت‌های تصمیم است. از این الگوریتم می‌توان جهت کلاسه‌بندی یا رگرسیون بهره برد، که در هر دو مورد در عمل، توانایی پیش‌بینی خوبی را از خود نشان داده است. به منظور کلاسه‌بندی یک نمونه جدید بر پایه ویژگی‌های آن نمونه، هر درخت، کلاسی که نمونه را متعلق به آن می‌داند مشخص می‌کند و در واقع یک رای می‌دهد. جنگل، بر اساس این آراء، دسته‌ای که بیشترین رای را به خود اختصاص داده انتخاب و به عنوان دسته نهایی نمونه داده تعیین می‌کند. مشابه این فرآیند برای عمل رگرسیون

بهتری نسبت به کرنل‌های دیگر (چند جمله‌ای و خطی) از خود نشان داده است و همچنین عملکرد این الگوریتم بسیار نزدیک به جنگل تصادفی می‌باشد. خروجی این دو الگوریتم از دید میانگین مطلق خطا تقریباً یکسان می‌باشد ولی جنگل تصادفی از دید ریشه میانگین مربعات خطا عملکرد بهتری داشته است. این تفاوت از آن جهت که معیار ریشه میانگین مربعات خطا جریمه بیشتری برای خطاهای بزرگتر در نظر می‌گیرد (اگر میزان خطای پیش‌بینی ۲ برابر شود با به توان ۲ رسیدن، جریمه آن ۴ برابر می‌شود) حاکی از آن است که پراکندگی خطای پیش‌بینی در الگوریتم جنگل تصادفی کمتر است. معیار دیگری که در جدول ۱ آمده است و نشان از برتری الگوریتم جنگل تصادفی دارد، نشان می‌دهد که چند درصد تغییرات متغیر هدف به وسیله متغیرهای مستقل تبیین می‌شود یا به عبارتی چه مقدار از واریانس موجود در متغیر هدف توسط الگوریتم قابل پیش‌بینی است. از این جنبه نیز جنگل تصادفی بهترین عملکرد را از خود نشان داده است.

دو کرنل دیگر به دلیل سادگی و عدم توانایی در مدلسازی پیچیدگی‌ها، نتایجی ضعیف‌تر از کرنل گوسی را نمایش می‌دهد.

جدول (۱): ارزیابی عملکرد الگوریتم‌ها

الگوریتم	مرحله	MAE	RMSE	R^2 score
جنگل تصادفی	آموزش	۰٫۲۸۳	۰٫۴۳۶	۰٫۶۸
	تست	۰٫۲۵۸	۰٫۳۵۰	
SVR با کرنل‌های گوسی	آموزش	۰٫۲۹۰	۰٫۴۷۴	۰٫۵۶
	تست	۰٫۲۵۹	۰٫۴۰۶	
SVR با کرنل‌های خطی	آموزش	۰٫۲۹۳	۰٫۴۷۷	۰٫۵۳
	تست	۰٫۲۶۶	۰٫۴۲۱	
SVR با کرنل‌های چندجمله‌ای	آموزش	۰٫۳۵۷	۰٫۵۲۶	۰٫۱۸
	تست	۰٫۳۸۱	۰٫۵۵۷	
رگرسیون خطی	آموزش	۰٫۲۹۸	۰٫۴۳۷	۰٫۵۲
	تست	۰٫۳۳۰	۰٫۴۲۵	

عملکرد ضعیف کرنل چند جمله‌ای نشان می‌دهد که انتقال داده‌ها به فضای جدید با استفاده از کرنل چندجمله‌ای منجر به تفکیک بهتری نشده و حتی کار را برای پیش‌بینی مشکل‌تر کرده است. ایجاد این پیچیدگی بیشتر منجر به عملکرد ضعیف‌تری حتی نسبت به رگرسیون خطی شده است.

واقعی و پیش‌بینی شده امکان‌پذیر نیست، به عنوان نمونه برای الگوریتم جنگل تصادفی این مقادیر برای ۳۰ نمونه از داده‌های دسته آزمون که بصورت تصادفی انتخاب شده‌اند در شکل ۲ رسم شده است. در این شکل محور عمودی مدت زمان تداوم وقفه به ساعت را نشان می‌دهد. محور افقی شماره نمونه را نمایش می‌دهد. خط آبی مقدار واقعی مدت زمان تداوم وقفه رخ داده و خط چین قرمز مقدار پیش‌بینی شده توسط الگوریتم مربوطه برای آن رخداد را نمایش می‌دهد. الگوریتم‌ها پیش‌بینی خوب و نزدیکی را در عمل نشان دادند.

این نزدیکی عملکرد و همچنین معیار متعارف ساعت برای مدت زمان تداوم وقفه، کار مقایسه را دشوار می‌کند. برای بررسی دقیق‌تر مقدار خطای هر الگوریتم در هر نمونه در شکل ۳ مقدار خطای الگوریتم‌ها در کنار هم رسم شده است. در این شکل خط چین قرمز مقدار واقعی مدت زمان تداوم وقفه و نقطه‌های رنگی مقدار مطلق خطای الگوریتم مربوطه را برای آن نمونه نمایش می‌دهد. هر چه مکان این نقطه پایین‌تر باشد، نمایانگر خطای کمتر آن الگوریتم در پیش‌بینی و متعاقباً عملکرد بهتر است. از بررسی این شکل مشهود است که نمی‌توان یک الگوریتم را پیروز قاطع این مسابقه دانست و نیاز به معدل‌گیری از عملکرد آن‌ها در فضای آزمون است. ازینرو برای ارزیابی دقیق عملکرد الگوریتم‌ها نیاز به معیاری جامع می‌باشد. در بخش بعد به مقایسه عملکرد روش‌ها پرداخته خواهد شد.

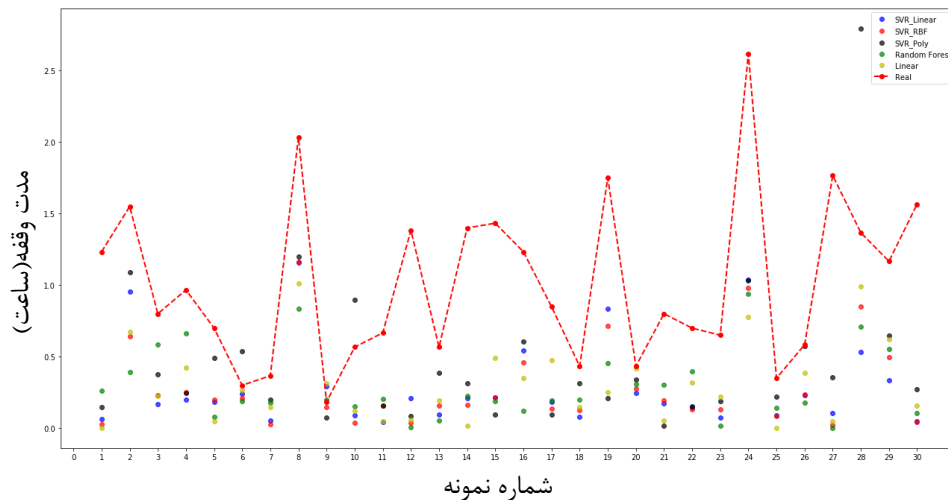
۴- ارزیابی عملکرد

برای ارزیابی عملکرد از سه معیار ریشه میانگین مربعات خطا ($RMSE^{(۴)}$)، میانگین مطلق خطا ($MAE^{(۵)}$) و ضریب تعیین R^2 یا $R^{(۴)}$ استفاده شده است. ضریب تعیین، به ما نشان می‌دهد که چه مقدار از واریانس موجود در متغیر هدف توسط مقداری که بواسطه متغیرهای مستقل پیش‌بینی شده، توصیف شده است [۳۴]. نتایج حاصل از این ارزیابی در جدول ۱ آورده شده است. همانطور که در نتایج مشاهده می‌شود الگوریتم جنگل تصادفی کمترین خطای پیش‌بینی را نسبت به باقی الگوریتم‌ها داشته است.

این الگوریتم توانسته است با میانگین مطلق خطای ۰٫۲۵۸ ساعت بطور نسبی بهترین پیش‌بینی را انجام دهد؛ به عبارتی این الگوریتم می‌تواند مدت زمان تداوم وقفه را با خطای متوسط ۱۵٫۴۸ دقیقه پیش‌بینی کند. در جایگاه‌های بعدی به ترتیب رگرسیون بردار پشتیبان با کرنل گوسی، رگرسیون بردار پشتیبان با کرنل خطی، رگرسیون خطی و رگرسیون بردار پشتیبان با کرنل چند جمله‌ای قرار دارند. کرنل گوسی به علت انعطاف‌پذیری بیشتری که دارد، عملکرد



شکل (۲): مقایسه مقدار واقعی و مقدار پیش‌بینی شده توسط الگوریتم جنگل تصادفی



شکل (۳): مقایسه میزان خطای پیش‌بینی الگوریتم‌ها

جدول (۲): زمان‌بندی و تعداد عملیات نگهداری و تعمیرات پیشگیرانه انجام شده در سال اول برای هر فیدر

شماره ماه	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
شماره فیدر	۱	۲	۰	۱	۰	۰	۰	۰	۱	۰	۰	۰
۲	۰	۰	۳	۰	۰	۰	۰	۰	۰	۱	۰	۰
۳	۱	۱	۰	۰	۰	۱	۱	۰	۰	۰	۰	۱
۴	۳	۲	۱	۰	۰	۰	۰	۱	۰	۰	۰	۱

ناخالصی به جهت تفکیک داده‌ها کمک کند. همانطور که در این شکل پیداست ویژگی مدت زمان رسیدن تیم نت کمک شایانی به این مدلسازی کرده است. پس از این ویژگی همانگونه که انتظار می‌رفت انرژی تامین نشده هم تاثیر بسزایی دارد. مورد قابل توجه بعدی تاثیر خوب پارامتر TLM بر روی مدلسازی و نتایج است. عبارتی این نشان از میزان تاثیرگذاری نت پیشگیرانه از مدت زمان تداوم وقفه دارد.

۵- تحلیل حساسیت زمان تداوم وقفه نسبت به عملیات نگهداری پیشگیرانه

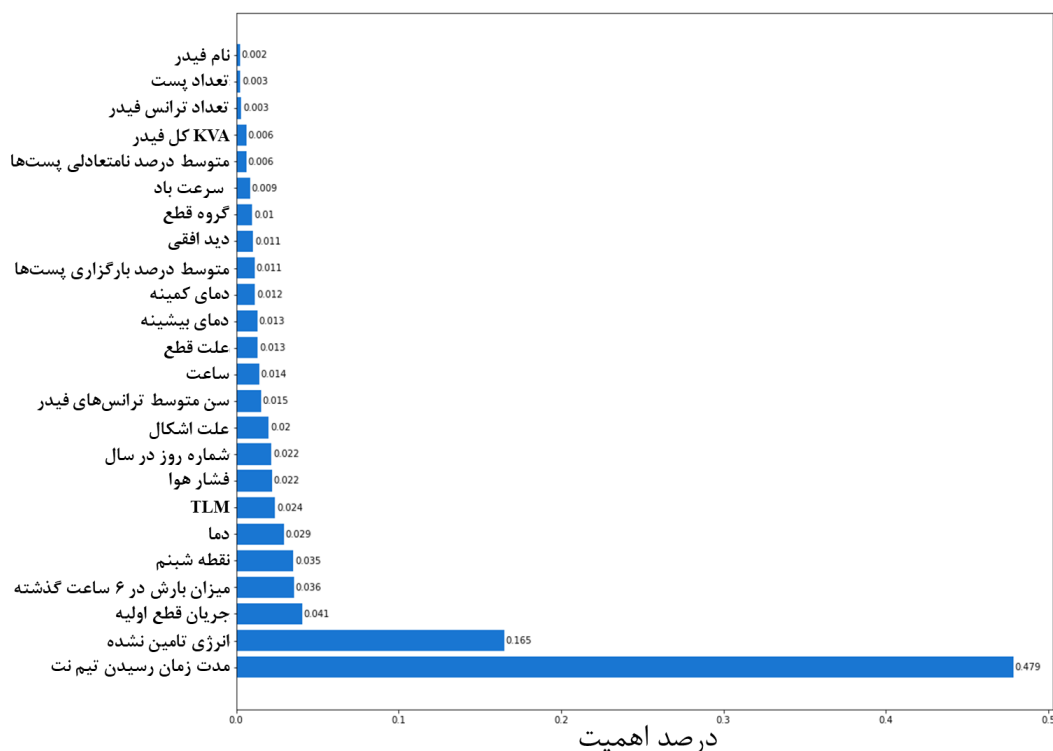
اثر عملیات نگهداری پیشگیرانه را می‌توان در دو بخش بررسی کرد. بخش اول پیشگیری و جلوگیری از بوجود آمدن برخی وقفه‌ها در شبکه است. بخش دوم اثریست که این نوع نگهداشت بر کاهش زمان تداوم وقفه‌های دیگر دارد، وقفه‌هایی که با وجود این نگهداشت نیز در شبکه رخ خواهند داد. عبارتی در این بخش هدف بررسی اثر نگهداشت

در مرجع [۲۸] کاری مشابه این پژوهش صورت پذیرفته است. در این مرجع با استفاده از علت قطع در کنار ۱۹ ویژگی دیگر به پیش‌بینی مدت زمان تداوم وقفه با استفاده از شبکه‌های عصبی بازگشتی پرداخته شده است. ریشه میانگین مربعات خطای پیش‌بینی در این مرجع ۴٫۰۲ ساعت است. شایان ذکر است که در ادامه کار در این مرجع با بهره‌گیری از گزارشات فرستاده شده توسط تیم تعمیرات این پیش‌بینی بروزرسانی می‌شود و خطای آن به ۳٫۹ کاهش می‌یابد. مقایسه خروجی این دو پژوهش بدلیل تفاوت در پایگاه داده شاید اقدام صحیحی نباشد، ولی می‌توان یکی از عوامل تاثیرگذار در بهبود نتایج را ویژگی‌های استخراج شده دانست.

در شکل ۴ درصد اهمیت هر یک از ویژگی‌های مستقل نمایش داده شده است. این درصد اهمیت از خروجی الگوریتم جنگل تصادفی استخراج شده است. مقدار بالاتر شاخص در نمودار نشان از اهمیت بیشتر هر ویژگی دارد. این شاخص که با عنوان اهمیت جینی^{۱۷} نیز شناخته می‌شود، به دنبال ویژگی می‌باشد که با بیشینه میزان کاهش

عملیات نگهداری پیشگیرانه در ۶ ماه اول سال نه تنها از وقوع برخی وقفه‌ها جلوگیری به عمل آمد، بلکه به میزان ۳۰ دقیقه در ۶ ماه اول سال برای ۴ فیدر مورد مطالعه، زمان بازبایی شبکه کاهش و رضایت مشترکین افزایش پیدا کرده است. شکل ۵ مدت زمان تداوم وقفه را در حضور عملیات نت پیشگیرانه و بدون آن برای ۸۵ وقفه ثبت شده در ۶ ماه اول سال نشان می‌دهد. خط قرمز که نشان دهنده عدم انجام نت پیشگیرانه است در اکثر موارد در بالای خط آبی قرار دارد که نشان از افزایش زمان تداوم وقفه است. افزایش ۳۰ دقیقه‌ای مدت زمان تداوم وقفه برآیند این افزایش‌ها و کاهش‌ها است. در شکل ۶ با قراردادن مدت زمان وقفه‌ها در ظرف‌های ۵ دقیقه‌ای و رسم نمودار هیستوگرام آن به بررسی دو سناریو پرداخته شده است. در این نمودار نیز می‌توان

پیشگیرانه بر نرخ تعمیر، و نه نرخ خرابی، است. برای بررسی این موضوع به حساسیت سنجی مدل منتخب، نسبت به تغییرات در پارامتر مرتبط با این عملیات، نیاز داریم. پارامتر مرتبط با زمان انجام عملیات نگهداری پیشگیرانه در مدل استخراج شده TLM می‌باشد. جدول ۲ زمان‌بندی و تعداد عملیات نت پیشگیرانه انجام شده در سال اول برای هر چهار فیدر مورد مطالعه را بطور خلاصه نشان می‌دهد. برای انجام این تحلیل از روش طراحی مطالعه مداخله‌ای^{۱۸} بهره برده شده است [۳۵]. این روش مطالعه برخلاف روش طراحی مطالعه مشاهده‌ای^{۱۹}، که گذشته‌نگر است و یا عبارتی گروه‌های مورد بررسی تنها بر اساس مشاهدات گذشته تفکیک می‌شوند، عمل می‌کند. روش طراحی مطالعه مداخله‌ای آینده‌نگر بوده و محقق با تشکیل چند



شکل (۴): درصد اهمیت ویژگی‌های ورودی

در کلیت شاهد بود که وقفه‌ها در حالت بدون نت پیشگیرانه از ظرف‌های کوچکتر به سمت ظرف‌های بزرگتر یا مدت زمان تداوم وقفه بیشتر مهاجرت کرده‌اند و عدم انجام نت پیشگیرانه به افزایش مدت زمان تداوم وقفه انجامیده است. نگاهی دیگر برای تحلیل خروجی بدست آمده بررسی تابع چگالی احتمال آن‌ها در دو حالت است. تخمین چگالی هسته^{۲۰} روشی غیرپارامتریک برای برآورد تابع چگالی احتمال یک متغیر تصادفی است. در این روش براساس تعداد محدودی از مشاهدات یا نمونه داده‌ها، استنباط‌هایی در رابطه با جامعه آماری آن انجام می‌شود. در شکل ۷ برای هر دو حالت مورد بررسی این تخمین صورت پذیرفته و تابع چگالی احتمال رسم شده است. با دقت در شکل می‌توان به این نکته رسید که با عدم انجام عملیات نت پیشگیرانه تابع

گروه به مطالعه اثر حضور پدیده‌ای می‌پردازد. در این مطالعه نیز، به کمک مشاهدات گذشته، رفتار آینده سیستم توسط مدل برازش شده در دو گروه حضور و عدم حضور نت پیشگیرانه مورد بررسی قرار گرفت. بدین منظور فرض گشت در ۶ ماه اول سال هیچ عملیات نت پیشگیرانه‌ای صورت نپذیرد. هدف بررسی میزان تغییرات مدت زمان تداوم وقفه در ۶ ماه اول سال ناشی از این تغییرات است. در ادامه ویژگی‌های ورودی به مدل، متناسب با فرض صورت پذیرفته اصلاح گشت. نتایج حاصله از شبیه‌سازی ۳۰ دقیقه افزایش مدت زمان تداوم وقفه برای این چهار فیدر در ۶ ماه اول سال را نشان می‌دهد. یادآوری می‌شود که این مدت زمان جدای وقفه‌هایی است که با انجام عملیات نت پیشگیرانه از وقوع آن پیشگیری شده است. عبارتی با انجام

جهت تحلیل و پیش‌پردازش انتخاب و استخراج شدند. در ادامه از روش‌های رگرسیون خطی و غیرخطی برای این پیش‌بینی استفاده شده است. این روش‌ها شامل رگرسیون خطی، رگرسیون بردار پشتیبان با کرنل‌های خطی، چند جمله‌ای، گوسی و همچنین جنگل تصادفی می‌شود.

نتایج پیش‌بینی نشانگر عملکرد بهتر روش‌های جنگل تصادفی و رگرسیون بردار پشتیبان با کرنل گوسی می‌باشد. الگوریتم جنگل تصادفی، به عنوان بهترین روش، قادر به پیش‌بینی مدت زمان تداوم وقفه با میانگین مطلق خطا ۱۵,۴۸ دقیقه می‌باشد. همچنین در انتها اثر زمان انجام عملیات نت پیشگیرانه بر مدت زمان تداوم وقفه بررسی گشت. از مدل بدست آمده می‌توان در پژوهش‌های آتی جهت پیش‌بینی مدت زمان تداوم وقفه، بهینه‌سازی عملیات نت پیشگیرانه و بهبود خدمات مشترکین بهره برد.

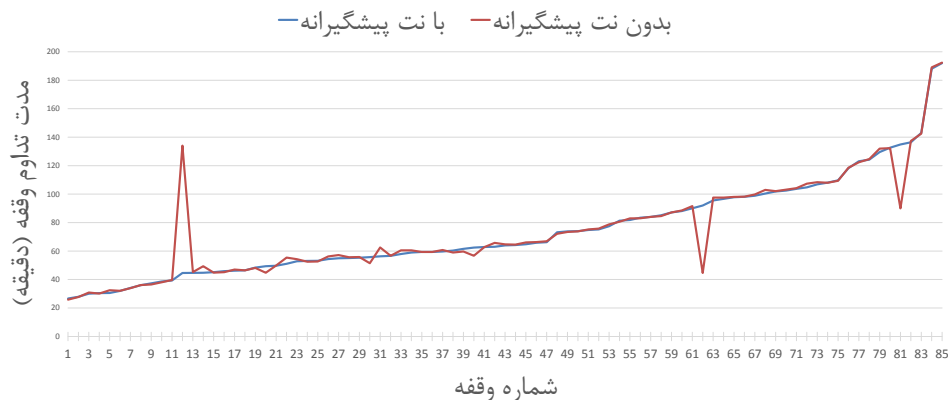
مراجع

- [1] W. Kersting, William H. Distribution system modeling and analysis. CRC press, 2017.
- [2] H. C. Caswell et al., "Weather normalization of reliability indices," IEEE trans. power deliv., vol. 26, no. 2, pp. 1273-1279, 2011.
- [3] A. Pahwa, M. Hopkins, and T. C. Gaunt, "Evaluation of

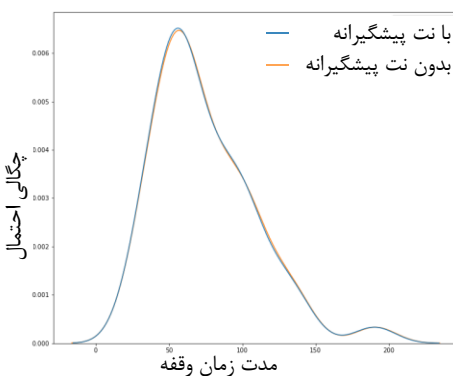
چگالی احتمال به سمت کاهش در میزان چگالی احتمال مد^{۲۱} و افزایش در دامنه راست خود پیشرفته است. عبارتی با عدم انجام نت پیشگیرانه با واریانس افزایش خواهد یافت و این افزایش پراکندگی رو به زمان‌های تداوم وقفه بیشتر خواهد بود.

۶- نتیجه‌گیری

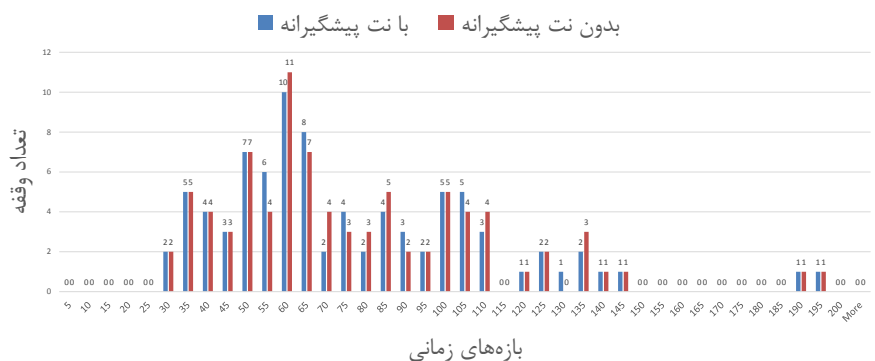
در این پژوهش، هدف مدلسازی مدت زمان تداوم وقفه با استفاده از اطلاعات استخراج شده از داشبورد بهره‌برداری شرکت برق و شرایط آب و هوایی است. بدین منظور در مرحله اول اطلاعات وقفه‌های شبکه در بازه زمانی ۴ ساله از پایگاه داده شرکت توزیع نیروی برق تهران بزرگ استخراج گردید. این رکوردها پس از پاکسازی با اطلاعات استخراج شده از سازمان هواشناسی تلفیق گردیده تا اطلاعات کاملی از شرایط موجود در دست باشد. این ۱۶ ویژگی شامل متوسط درصد بارگزاری پست‌ها، KVA کل ترانس‌های فیدر، متوسط درصد نامتعادلی پست‌ها، سن متوسط ترانس‌های فیدر، ساعت وقفه، مدت زمان رسیدن تیم نت، TLM، علت قطع، علت اشکال، گروه قطع، انرژی تامین نشده، تعداد پست فیدر، تعداد ترانس فیدر، جریان قطع اولیه، شماره روز در سال و نام فیدر فشار متوسط می‌شود. این ویژگی‌ها به همراه ۸ ویژگی‌های آب و هوایی به عنوان ورودی‌های اولیه



شکل (۵): مدت زمان تداوم وقفه با و بدون انجام عملیات نت پیشگیرانه



شکل (۷): تابع چگالی احتمال تخمینی



شکل (۶): هیستوگرام مدت زمان تداوم وقفه، با و بدون انجام عملیات نت پیشگیرانه

- [20] E. Akhavan-Rezai, M.-R. Haghifam, and A. Fereidunian, "Adaptive network-based fuzzy inference for momentary failure rate modeling," in IEEE PES T&D 2010, 2010, pp. 1–6.
- [21] M.-R. Haghifam, E. Akhavan-Rezai, and A. Fereidunian, "Failure rate modeling: A non-parametric data mining approach to MV network field data," in 2009 IEEE Electrical Power & Energy Conference (EPEC), 2009, pp. 1–6.
- [22] M. Gholami, M. Moeini-Aghtaie, A. Safdarian, and A. A. T. Fard, "Maintenance based distribution network expansion planning," 2017 Iranian Conference on Electrical Engineering (ICEE), 2017.
- [23] D. Wanik, E. Anagnostou, B. Hartman, and T. Layton, "Estimated time of restoration (ETR) guidance for Electric distribution networks," J. Homel. Secur. Emerg. Manag., vol. 15, no. 1, 2018.
- [24] A. Arif, S. Ma, Z. Wang, J. Wang, S. M. Ryan, and C. Chen, "Optimizing service restoration in distribution systems with uncertain repair time and demand," IEEE Trans. Power Syst., vol. 33, no. 6, pp. 6828–6838, 2018.
- [25] M.-Y. Chow, L. S. Taylor, and M.-S. Chow, "Time of outage restoration analysis in distribution systems," IEEE trans. power deliv., vol. 11, no. 3, pp. 1652–1658, 1996.
- [26] M. M. Adibi and D. P. Milanicz, "Estimating restoration duration," IEEE Trans. Power Syst., vol. 14, no. 4, pp. 1493–1498, 1999.
- [27] J. R. A. Rodriguez and A. Vargas, "Fuzzy-heuristic methodology to estimate the load restoration time in MV networks," IEEE Trans. Power Syst., vol. 20, no. 2, pp. 1095–1102, 2005.
- [28] A. Jaech, B. Zhang, M. Ostendorf, and D. S. Kirschen, "Real-time prediction of the duration of distribution system outages," IEEE Trans. Power Syst., vol. 34, no. 1, pp. 773–781, 2019.
- [۲۹] ن. سالک گیلانی، ع. فریدونیان، "پیش‌بینی مدت زمان وقفه در شبکه توزیع با داده‌کاوی"، هشتمین کنفرانس منطقه‌ای سیرد، اسفند ماه ۱۳۹۹، توانیر، ایران.
- [30] A. Burkov, The Hundred-Page Machine Learning Book. USA: Creative Commons., 2019.
- [31] C.-C. Chang and C.-J. Lin, "Libsvm," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1–27, 2011.
- [32] S. Fattaheian-Dehkordi, A. Fereidunian, H. Gholami-Dehkordi, and H. Lesani, "Hour-ahead demand forecasting in smart grid using support vector regression (SVR)," International Transactions on Electrical Energy Systems, vol. 24, no. 12, pp. 1650–1663, 2013.
- [33] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," Stat. Comput., vol. 27, no. 3, pp. 659–678, 2017.
- [34] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning: With applications in R, 1st ed. New York, NY: Springer, 2013.
- [35] M. S. Thiese, "Observational and interventional study design types; an overview," Biochemia Medica, vol. 24, no. 2, pp. 199–210, 2014.
- outages in overhead distribution systems of south africa and of Manhattan," 2007.
- [4] P. Kankanala, S. Das, and A. Pahwa, "AdaBoost+: An ensemble learning approach for estimating weather-related outages in distribution systems," IEEE Trans. Power Syst., vol. 29, no. 1, pp. 359–367, 2014.
- [5] A. Bostani, S. M. Moghimi, R. Dashti, and S. M. Hashemi, "The Role of Preventive Major Maintenance in the Costs of Electric Energy Distribution Companies," Indian Journal of Science and Technology, vol. 11, no. 19, pp. 1–7, 2018.
- [6] M. Pourahmad, I. Ahmadi, and T. Barforoushi, "Modeling of Impacts of Loading on Transformer Reliability," Journal of Iranian Association of Electrical and Electronics Engineers, vol. 14, no. 1, pp. 63–72, 2017.
- [7] T. Sami, SM. Shahrtash, and A. Gholami, "High Voltage Power Cable Life Estimation under Thermal, Electrical and Multi-Stresses Conditions," Journal of Iranian Association of Electrical and Electronics Engineers, vol. 7, no. 1, pp. 61–66, 2010.
- [8] P. T. Baboli, D. Babazadeh, A. Raeiszadeh, S. Horodyvsky, and I. Koprek, "Optimal Temperature-Based Condition Monitoring System for Wind Turbines," Infrastructures, vol. 6, no. 4, p. 50, 2021.
- [9] M. Mollahassani-pour, et al. "Long Term Cost Effective Preventive Maintenance Scheduling," Journal of Iranian Association of Electrical and Electronics Engineers 12.3 (2016): 49-54.
- [10] J. V. Casteren, M. Bollen, and M. Schmieg, "Reliability assessment in electrical power systems: the Weibull-Markov stochastic model," IEEE Transactions on Industry Applications, vol. 36, no. 3, pp. 911–915, 2000.
- [11] J. Van Casteren, "Assessment of interruption costs in electric power systems using the weibull-Markov model," Chalmers University of Technology, 2003.
- [12] N. Cross, R. Herman, and C. T. Gaunt, "Investigating the usefulness of the Beta pdf to describe Parameters in Reliability Analyses," in 2006 International Conference on Probabilistic Methods Applied to Power Systems, 2006, pp. 1–6.
- [13] T. H. Ortmeier and B. M. Fisk, "Characterization of distribution system interruption duration," in 2012 IEEE Power and Energy Society General Meeting, 2012, pp. 1–5.
- [۱۴] ن. سالک گیلانی، ع. فریدونیان، "مدلسازی داده رانه مدت زمان تداوم وقفه در سیستم توزیع انرژی الکتریکی"، دهمین کنفرانس شبکه های هوشمند انرژی، کاشان، آذر ۱۳۹۹.
- [15] D. Zhu, D. Cheng, R. P. Broadwater, and C. Scirbona, "Storm modeling for prediction of power distribution system outages," Electric Power Syst. Res., vol. 77, no. 8, pp. 973–979, 2007.
- [16] Y. Zhou, A. Pahwa, and S.-S. Yang, "Modeling weather-related failures of overhead distribution lines," IEEE Trans. Power Syst., vol. 21, no. 4, pp. 1683–1690, 2006.
- [17] H. Liu, R. A. Davidson, and T. V. Apanasovich, "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms," Reliab. Eng. Syst. Saf., vol. 93, no. 6, pp. 897–912, 2008.
- [18] K. Alvehag and L. Soder, "A reliability model for distribution systems incorporating seasonal variations in severe weather," IEEE trans. power deliv., vol. 26, no. 2, pp. 910–919, 2011.
- [19] A. Domijan et al., "Effects of normal weather conditions on interruptions in distribution systems," Int. J. Power Energy Syst., vol. 25, no. 1, 2005.

زیر نویس‌ها

- ¹ Generalized Extreme Value distribution
- ² Time since last maintenance action (TLM)
- ³ Missing data
- ⁴ Support Vector Regression
- ⁵ Random Forest
- ⁶ Support Vector Machine
- ⁷ Grid search
- ⁸ Breiman
- ⁹ Imbalanced problems
- ¹⁰ K-fold Cross Validation
- ¹¹ Scale
- ¹² Standardization
- ¹³ Normalization
- ¹⁴ Root Mean Square Error
- ¹⁵ Mean Absolute Error
- ¹⁶ Coefficient of Determination
- ¹⁷ Gini Importance
- ¹⁸ Interventional study design
- ¹⁹ Observational study design
- ²⁰ Kernel Density Estimation (KDE)
- ²¹ Mode