



Computational Intelligence in Electrical Engineering
Vol. 16, No. 4, 2026
pp. 43-58
Research Paper

Outage Cause Identification in Electric Power Distribution Systems Using an Ensemble Learning-Based Method and Logistic Regression

Mohammad Reza Dehbozorgi¹, Alireza Hamedei², Mohammad Ali Nematollahi³

¹ School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

² Assistant Professor, Department of Electrical Engineering, Faculty of Engineering, Fasa University, Fasa, Iran

³ Assistant Professor, Department of Computer Sciences, Fasa University, Fasa, Iran

Abstract:

In electric power distribution systems, reliability is of paramount importance, and the occurrence of unscheduled outages affects customer satisfaction and operational costs. Detecting the causes of permanent and momentary outages, particularly in grids with limited observability, is time-consuming and heavily dependent on field operations and experience. This paper proposes a data-driven framework based on ensemble learning and logistic regression. The proposed approach leverages more than 50,000 records of unscheduled outages from a real-world distribution system, along with hot-line and cold-line maintenance records, hourly feeder load data, and weather data, to extract rich descriptive features. Using a One-vs-Rest strategy combined with resampling, we address the severe class imbalance in outage data. Ensemble models are trained for each outage cause and are then combined via soft voting, with weights determined by validation AUC. This prioritizes the most likely causes of permanent and momentary outages. The results demonstrate high classification accuracy across different outage causes, achieving up to a 30% improvement in the accuracy of identifying different outage causes compared to state-of-the-art methods. Overall, the proposed approach enhances the distribution system's reliability by reducing outage duration and accelerating outage cause identification.

Keywords: Distribution Network Reliability, Ensemble Learning, Logistic Regression, Permanent and Momentary Outages, Power Distribution Network.



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/isee.2026.148092.1778>

شناسایی علت خاموشی‌های رخ داده در سیستم توزیع انرژی الکتریکی با استفاده از روش

مبتنی بر یادگیری گروهی و رگرسیون لجستیک

محمدرضا ده بزرگی^۱، علیرضا حامدی^{۲*}، محمدعلی نعمت‌اللهی^۳

۱- دانش آموخته دکتری، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز، ایران

m.dehbozorgi@shirazu.ac.ir

۲- استادیار گروه برق، دانشکده مهندسی، دانشگاه فسا، فسا، ایران

ar-hamedi@fasau.ac.ir

۳- استادیار گروه علوم کامپیوتر، دانشکده علوم پایه، دانشگاه فسا، فسا، ایران

ma.nematollahi@fasau.ac.ir

چکیده: در سیستم‌های توزیع انرژی الکتریکی، قابلیت اطمینان اهمیت زیادی دارد و وقوع خاموشی‌های برنامه‌ریزی نشده بر رضایت مشترکین و هزینه‌های بهره‌برداری اثر می‌گذارد. شناسایی علل خاموشی‌های برنامه‌ریزی نشده دائم و گذرا، به ویژه در شبکه‌های با رویت‌پذیری محدود، زمان‌بر و وابسته به عملیات میدانی است. این مقاله چارچوبی داده‌محور مبتنی بر یادگیری گروهی و رگرسیون لجستیک پیشنهاد می‌دهد که با بهره‌گیری بالغ بر ۵۰ هزار رکورد واقعی خاموشی برنامه‌ریزی نشده، سوابق تعمیرات خط سرد و گرم، بار ساعتی فیدرها و اطلاعات آب‌وهوایی، ویژگی‌هایی غنی استخراج می‌کند. علاوه بر آن، با رویکرد One-vs-Rest و نمونه‌برداری مجدد برای مدیریت نامتعادلی شدید کلاس‌های موجود در مجموعه داده خاموشی، مدل‌های گروهی برای هر علت خاموشی آموزش دیده و با رأی‌گیری نرم و وزن‌دهی بر اساس AUC، به دست آمده از مجموعه اعتبارسنجی، ترکیب می‌شوند تا علل محتمل خاموشی‌های دائم و گذرا اولویت‌بندی شوند. نتایج نشان‌دهنده دقت زیاد طبقه‌بندی علل مختلف خاموشی است که دقت شناسایی علل مختلف را تا ۳۰ درصد نسبت به روش‌های قبلی بهبود می‌دهد. مجموعه این روش‌ها باعث افزایش قابلیت اطمینان شبکه از طریق کاهش زمان خاموشی و تسریع در یافتن علت خاموشی می‌شود.

واژه‌های کلیدی: خاموشی‌های دائم و گذرا، رگرسیون لجستیک، شبکه توزیع انرژی الکتریکی، قابلیت اطمینان شبکه توزیع، یادگیری گروهی.

۱- مقدمه

برقراری سرویس را با چالش مواجه کند [۱]. محدودیت رویت‌پذیری در بسیاری از سیستم‌های توزیع، به ویژه در فیدرهایی با سطح نفوذ پایین تجهیزات و ادوات هوشمند نیز سبب می‌شود شناسایی علت خاموشی و تعیین تجهیز معیوب زمان‌بر و متکی بر عملیات میدانی باشد [۲]. از این رو، کاهش مدت و فراوانی خاموشی‌ها و بهبود شاخص‌های قابلیت اطمینان مانند SAIDI^۱ و SAIFI^۲، مستلزم درکی دقیق‌تر از الگوهای وقوع خاموشی‌های شبکه توزیع و عوامل مؤثر بر آنهاست [۳].

قابلیت اطمینان سیستم‌های توزیع انرژی الکتریکی، با توجه به اتصال مستقیم مشترکین به آنها، یکی از مهم‌ترین شاخص‌های عملکرد این سیستم‌ها محسوب می‌شود و به طور مستقیم بر رضایت مشترکین، هزینه‌ها و الزامات بهره‌برداری تأثیر می‌گذارد. با توجه به ساختار عمدتاً شعاعی فیدرهای توزیع، وقوع خطا ممکن است به خاموشی بخشی قابل ملاحظه از مشترکین منجر شود و فرایند یافتن محل خطا و

بر نرخ وقوع خاموشی‌ها بررسی شده و در [۱۰]، با یک رهیافت آماری، تعداد خاموشی‌های ناشی از رعدوبرق به صورت روزانه پیش‌بینی شده است.

در مقابل، دسته‌ای دیگر از مطالعات بر شناسایی علت خاموشی‌های رخ داده برای تعمیر سریع اصلاحی یا استخراج عوامل مؤثر بر وقوع آنها به منظور پشتیبانی از برنامه‌های نگهداری پیشگیرانه تمرکز دارند. برخی از پژوهش‌ها با تکیه بر داده‌های شکل موج خطا، داده‌های حسگرهای پیشرفته مانند PMUهای سطح توزیع و شکل موج‌های خطا و مدل‌های طبقه‌بندی، علت خاموشی را در زمان وقوع تشخیص می‌دهند [۱۱، ۱۲]؛ اما محدودیت دسترسی به این داده‌ها با رزولوشن زمانی زیاد در بسیاری از فیدرهای توزیع، بسیاری از مطالعات را به سمت داده‌های عملیاتی عمومی‌تر مانند سوابق خاموشی‌های برنامه‌ریزی نشده سوق داده است. در این میان، برخی از پژوهش‌ها از قواعد انجمنی^۳ برای شناسایی عوامل مؤثر بر خاموشی‌ها بهره گرفته‌اند. برای مثال، [۱۳] صرفاً بر خاموشی‌های با منشأ حیوانات و پوشش گیاهی تمرکز داشته و [۵] با لحاظ بار ایستگاه فوق توزیع، علاوه بر علل طبیعی، به خطاهای تجهیز محور نیز رسیدگی کرده است. همچنین، در [۱۴، ۱۵]، با استفاده از مدل‌های یادگیری ماشین مانند درخت تصمیم و جنگل تصادفی، علت خاموشی از طریق آموزش یک مدل طبقه‌بندی پیش‌بینی شده است. [۱] نیز از روش‌هایی مانند LR و شبکه عصبی مصنوعی (ANN)^۴ برای تشخیص دو علت پوشش گیاهی و حیوانات بهره برده است.

با وجود تلاش‌های انجام‌شده پیشین، بیشتر مطالعات به مجموعه‌ای محدود از علل خاموشی (معمولاً علل طبیعی مانند حیوانات و پوشش گیاهی) محدود شده‌اند یا تجهیزات شبکه را به صورت کلی در نظر گرفته‌اند و تفاوت رفتار تجهیزات مختلف کمتر بررسی شده است. همچنین، بهره‌گیری هم‌زمان از داده‌های خاموشی و سوابق کامل تعمیرات و نگهداری (به ویژه تعمیرات خط سرد و گرم) در بیشتر این پژوهش‌ها مشاهده نمی‌شود. افزون بر این، به جز مرجع [۱۶] که صرفاً به خاموشی‌های گذرا اختصاص یافته، شناسایی علل خاموشی‌های دائم و گذرا در قالب یک چارچوب داده‌محور

خاموشی‌های سیستم توزیع به طور کلی به دو دسته خاموشی‌های دائم و خاموشی‌های گذرا تقسیم می‌شوند. خاموشی‌های دائم ناشی از خرابی یا خطای پایدار تجهیزات و مستلزم عیب‌یابی میدانی، تعمیر یا تعویض اجزا و ثبت دقیق اطلاعات در سیستم‌های بهره‌برداری هستند، در حالی که خاموشی‌های گذرا ماهیتی موقتی دارند و معمولاً از طریق بازست خودکار یا دستی مدار برطرف می‌شوند [۴]. تشخیص صحیح علت زمینه‌ای خاموشی (تجهیز خطادیده یا علل طبیعی مانند برخورد پرندگان و حیوانات یا پوشش گیاهی)، نقشی اساسی در کاهش مدت و فراوانی خاموشی‌ها و طراحی راهبردهای مؤثر تعمیرات اصلاحی دارد؛ به گونه‌ای که آگاهی اولیه از علت احتمالی خاموشی‌های دائم می‌تواند فرایند یافتن محل خطادیده و برقراری سرویس را تسریع کند [۵]. از سوی دیگر، علت خاموشی‌های گذرا معمولاً به دلیل عدم اجرای تعمیرات اصلاحی ثبت نمی‌شود، در حالی که تحلیل این رخ داده‌ها و تشخیص علت واقعی آنها ممکن است به شناسایی عوامل زوال شبکه و بهبود شاخص‌هایی مانند SAIFI منجر شود. به نظر می‌رسد بهره‌گیری از تحلیل‌های داده‌محور روی مجموعه داده‌های متنوع موجود در شبکه توزیع شامل سوابق خاموشی‌های برنامه‌ریزی شده و برنامه‌ریزی نشده، سوابق تعمیرات و نگهداری تجهیزات و سایر داده‌های بهره‌برداری، امکان استخراج الگوهای پنهان حاکم بر خاموشی‌ها و ارتقای شاخص‌های قابلیت اطمینان سیستم‌های توزیع را فراهم می‌کند. در این راستا، مطالعات اخیر از رویکردهای یادگیری ماشین و یادگیری گروهی برای بهبود مدیریت و کاهش تأثیر خاموشی‌ها در سیستم‌های توزیع بهره برده‌اند [۶، ۷].

به این منظور، در سال‌های اخیر و با گسترش روش‌های داده‌کاوی و یادگیری ماشین، استفاده از این رویکردها برای تحلیل خاموشی‌های رخ داده در سیستم‌های توزیع و بهره‌برداری از دانش استخراج شده برای افزایش قابلیت اطمینان شبکه مورد توجه قرار گرفته است. برای مثال، در [۸]، تعداد خاموشی‌های ناشی از پوشش گیاهی و رعدوبرق به صورت ماهانه و با استفاده از مدل‌های سری زمانی پیش‌بینی شده است. همچنین، در [۹]، اثر درخت‌زنی و هرس پوشش گیاهی

یکپارچه مورد توجه قرار نگرفته است.

در این مقاله، به منظور برطرف کردن کاستی‌های شناسایی شده در پژوهش‌های پیشین، یک چارچوب داده‌محور مبتنی بر یادگیری گروهی و رگرسیون لجستیک (LR)^۵ برای شناسایی علل احتمالی خاموشی‌های دائم و گذار در سیستم توزیع ارائه می‌شود. یکی از چالش‌های اصلی در این مسأله نامتعادلی شدید داده‌هاست؛ به گونه‌ای که برخی از علل خاموشی فراوانی بسیار بیشتری نسبت به سایر علل دارند. برای مواجهه با این چالش، از یک رویکرد یادگیری گروهی مبتنی بر نمونه‌برداری مجدد^۶ استفاده می‌شود. در این رویکرد، برای هر علت خاموشی، چندین زیرمجموعه از داده‌ها با استفاده از ضرایب نمونه‌برداری متفاوت (α) ایجاد می‌شوند تا آن علت به صورت متناوب در موقعیت کلاس اکثریت یا اقلیت قرار گیرد. سپس، برای هر زیرمجموعه، یک مدل رگرسیون لجستیک آموزش دیده می‌شود و در نهایت، این مدل‌ها از طریق رأی‌گیری نرم^۷ و وزندهی بر اساس AUC^8 مجموعه اعتبارسنجی با یکدیگر ترکیب می‌شوند. این سازوکار اجازه می‌دهد مدل گروهی تنوع کامل رکوردها را مشاهده و به ویژه الگوهای حاکم بر کلاس‌های اقلیت را با دقت و پایداری بیشتری شناسایی کند. خروجی این چارچوب به صورت وزندهی شده برای شناسایی علل احتمالی خاموشی‌های دائم و گذار به کار گرفته می‌شود. ضمناً، برای افزایش غنای ویژگی‌های ورودی و بهبود دقت طبقه‌بندی، در این پژوهش، از مجموعه داده‌های شبکه توزیع شامل خاموشی‌های برنامه‌ریزی نشده و برنامه‌ریزی شده، سوابق تعمیرات و نگهداری به روش‌های سرد و گرم، بار فیدرهای توزیع و اطلاعات آب‌وهوایی استفاده شده است. ترکیب این منابع داده امکان تحلیل یکپارچه رفتار شبکه و ارتباط بین وضعیت بهره‌برداری، شرایط محیطی و عملکرد تجهیزات را فراهم می‌کند. به طور خلاصه، نوآوری‌های اصلی این پژوهش به شرح زیر هستند:

- ارائه چارچوب یکپارچه مبتنی بر یادگیری گروهی و رگرسیون لجستیک برای شناسایی هم‌زمان علل خاموشی‌های دائم و گذار، در حالی که مطالعات قبلی فقط یکی از دو نوع

خاموشی را بررسی کرده‌اند.

- بهره‌گیری هم‌زمان از داده‌های خاموشی‌های برنامه‌ریزی شده، سوابق تعمیرات و نگهداری به روش‌های سرد و گرم و بار فیدرهای توزیع به منظور استخراج ویژگی‌های توصیف‌کننده و ارتقای دقت مدل شناسایی علل خاموشی‌ها.

- مدیریت مؤثر نامتعادلی شدید کلاس‌ها با ترکیب راهبرد One-vs-Rest، نمونه‌برداری مجدد با ضرایب مختلف و ترکیب مدل‌های گروهی از طریق رأی‌گیری نرم وزن‌دار بر اساس AUC اعتبارسنجی و ارائه خروجی به صورت اولویت‌بندی احتمالاتی علل محتمل که تصمیم‌گیری عملی و سریع برای بهره‌برداران را تسهیل می‌کند.

در ادامه این مقاله، ابتدا در بخش دوم، مجموعه داده‌های استفاده شده و ویژگی‌های استخراج شده از آنها بررسی می‌شوند. سپس، بخش سوم و چهارم به ارائه مبانی روش پیشنهادی مبتنی بر مدل‌های گروهی و LR برای شناسایی خاموشی‌های دائم و گذار اختصاص دارند. در بخش پنجم، نتایج سه مطالعه موردی ارائه می‌شود و در نهایت، بخش ششم به جمع‌بندی و بیان مسیرهای پژوهشی آینده اختصاص دارد.

۲- مجموعه داده‌های استفاده شده

این بخش به معرفی مجموعه داده‌های پژوهش (شامل خاموشی‌های برنامه‌ریزی شده و برنامه‌ریزی نشده، بار فیدرها، سوابق تعمیراتی و اطلاعات آب و هوایی)، مراحل پیش‌پردازش و استخراج ویژگی برای مدل‌های یادگیری ماشین اختصاص دارد.

۲-۱- خاموشی‌های برنامه‌ریزی نشده

مجموعه داده خاموشی‌های برنامه‌ریزی نشده شامل ۵۰,۸۸۸ رکورد از سال‌های ۱۳۹۹ تا ۱۴۰۳ در ۲۴ منطقه عملیاتی شبکه توزیع مورد بررسی است و هسته اصلی داده‌های پژوهش را تشکیل می‌دهد. این داده‌ها که در زمان وقوع خاموشی‌های سطح فشار متوسط ثبت می‌شوند، عناوین کلیدی مانند زمان و تاریخ وقوع، شناسه فیدر، امور مربوط، نام پست فوق توزیع،

عملیاتی جایگزین شدند.

۲-۲- خاموشی‌های برنامه‌ریزی‌شده، تعمیرات خط

سرد و خط گرم

مجموعه داده خاموشی‌های برنامه‌ریزی‌شده شامل ۱۹,۲۷۸ رکورد از ۱۳۹۹ تا ۱۴۰۳ است که عمدتاً برای تعمیرات پیشگیرانه، تعویض تجهیزات فرسوده یا ارتقای شبکه ثبت شده‌اند. تحلیل‌ها نشان داد علل پرتکرار شامل سرویس و آچارکشی (۱۶ درصد، عمدتاً پست‌های زمینی/هوایی)، تعلیل/تعویض ترانسفورماتور (۶ درصد) و تعمیر تابلوی فشارضعیف (۲ درصد) است. برچسب‌ها با استفاده از تجهیزات تعمیرشده یا نوع تعمیر مانند پست/ترانسفورماتور، ادوات کلیدزنی، کابل، پایه، جمپر، خط هوایی، سرکابل، برق‌گیر، شاخه‌زنی و تعمیرات عمومی تجمیع شدند. از این داده، ویژگی‌های کلیدی مانند «فاصله زمانی تا آخرین تعمیر در تجهیز X» (مانند پست، ترانسفورماتور، کابل، جمپر، مقره یا درخت‌زنی) استخراج شدند که تعداد روزهای گذشته از تعمیر اخیر در فیدر دارای خاموشی را نشان می‌دهد.

مجموعه تعمیرات خط گرم (۹,۶۶۹ عملیات) عمدتاً شامل درخت‌بری (۳۲ درصد)، تعویض جمپر (۱۸ درصد) و نصب کاور سوزنی (۱۵ درصد) است. تعمیرات خط سرد (۲۸,۳۶۴ عملیات) نیز عمدتاً به تعویض تجهیزات معیوب (۴۵ درصد)، نصب ادوات حفاظتی (۳۰ درصد) و اقدامات اصلاحی پایه‌ها (۲۵ درصد) اختصاص دارد. از اینها نیز ویژگی‌های مشابه (فاصله تا تعمیر در تجهیزات مختلف مانند کابل/اتصالات، مقره‌ها، پوشش گیاهی، ادوات کلیدزنی، پایه/ایراق، جمپر و پست) استخراج شدند. مقادیر گمشده با میانگین امور پر و به نزدیک‌ترین عدد صحیح گرد شدند تا واحد روز حفظ شود.

این ویژگی‌ها برای ارزیابی اثربخشی تعمیرات پیشین مفید هستند. برای مثال، اگر خاموشی کوتاهی پس از تعمیر رخ دهد، ممکن است نشان‌دهنده نقص در تعمیر باشد، در حالی که فاصله طولانی ممکن است نشانه فرسودگی طبیعی تجهیزات و عدم توجه به تعمیرات پیشگیرانه باشد.

جریان قطع‌شده، محل قطع فیدر و علت خاموشی را پوشش می‌دهند. علت خاموشی، به عنوان متغیر هدف، نقش مرکزی دارد. پیش‌پردازش با هدف آماده‌سازی برای مدل‌های یادگیری ماشین شامل حذف عناوین غیرقابل استفاده در لحظه وقوع مانند توضیحات حادثه، مدت خاموشی، ENS^۹ و محل دقیق وقوع حادثه است. همچنین، برچسب‌های علت (شامل بیش از ۸۶ نوع) بر اساس دستورالعمل ثبت حوادث توانیر تجمیع شدند [۱۷]. این فرایند تعداد کلاس‌ها را به ۲۰ کاهش داد و نامتعادلی را تا حدی تعدیل کرد. جدول (۱) سهم علل و مقدار درصد ENS ناشی از آنها را نشان می‌دهد.

همان‌طور که در جدول (۱) مشاهده می‌شود، عملکرد یا نقص در کات‌اوت فیوز (۷۳ درصد)، خاموشی‌های گذرا (۱۱/۶ درصد) و حیوانات (۲/۴ درصد) مهم‌ترین علل خاموشی هستند. تحلیل‌ها حاکی از آن است که بیشتر خاموشی‌های کات‌اوت فیوز ناشی از عوامل خارجی مانند برخورد پرنده یا درخت است، نه خرابی ذاتی فیوز و با وجود سهم بالا در تعداد، ENS ناچیزی دارند؛ زیرا معمولاً فقط یک پست توزیع یا شاخه جانبی (T-Off) را قطع می‌کنند. در مقابل، خاموشی‌های گذرا با قطع کل فیدر، با وجود آنکه مدت‌زمان وقوعشان کوتاه و نیز، تعدادشان از خاموشی‌های ناشی از «عملکرد یا اشکال در کات‌اوت فیوز» کمتر است، سهمی قابل ملاحظه از ENS را به خود اختصاص می‌دهند.

برای آموزش مدل، ویژگی‌های زمانی (شامل فصل و ساعت وقوع به منظور بررسی تأثیر شرایط جوی و پیک بار)، ویژگی‌های مکانی (منطقه عملیاتی برای تحلیل تفاوت‌های جغرافیایی) و محل قطع مدار (قطعی یا عدم آن از ابتدای فیدر) استخراج شدند. افزون بر این، از داده‌های خاموشی‌های برنامه‌ریزی‌نشده، دو ویژگی مرتبط با خاموشی‌های گذرا شامل فاصله زمانی تا آخرین خاموشی گذرا و تعداد خاموشی‌های گذرا در ۳۰ روز گذشته در همان فیدر استخراج شدند که به ترتیب نشان‌دهنده احتمال تکرار کوتاه‌مدت و وجود مشکلات مزمن در فیدر هستند. مقادیر ناموجود ابتدا با میانگین همان فیدر در سایر رخ داده‌ها و در صورت نبود، با میانگین منطقه

جدول (۱): فرکانس وقوع خاموشی‌ها و سهم ENS ناشی از وقوع خاموشی‌ها با علل مختلف در سال‌های ۱۳۹۹-۱۴۰۳

علت خاموشی	درصد وقوع	ENS (مگاوات-ساعت)	علت خاموشی	درصد وقوع	ENS (مگاوات-ساعت)
ادوات کلیدزنی (بریکر و سکسیونر)	۱/۰۳	۵۱۷/۶	دیگر علل خاموشی	۰/۲۴	۳۶/۱
برق‌گیر	۰/۳۳	۱۵۸/۱	سرکابل	۱/۳۰	۸۵۶/۳
پایه	۰/۴۷	۳۴۵/۲	شرایط بد آب‌وهوایی	۱/۳۵	۴۴۴/۷
پوشش گیاهی	۰/۳۵	۱۸۹/۹	عملکرد یا اشکال در کات اوت فیوز	۷۲/۸۱	۷۹۹/۳
تجهیزات پست توزیع	۱/۰۷	۳۷/۴	عوامل خارجی (انسانی و متفرقه)	۱/۰۶	۳۲۴/۳
ترانسفورماتور	۰/۷۱	۲۲/۴	عیب در تجهیزات فوق توزیع	۰/۰۳	۱۵/۶
چمبر	۱/۳۴	۳۱۲/۳	عیب نامشخص	۱/۱۵	۷۰۶/۱
حیوانات	۲/۴	۱۳۹/۶	کابل	۱/۰۲	۶۷۴
خاموشی گذرا	۱۱/۶۳	۱۰۴۹/۶	مشکلات سمت مشترکین	۰/۷۵	۳۸۷/۸
خط توزیع هوایی و متعلقات آن	۰/۵۸	۳۷۸	مقره	۰/۴	۲۴۶/۵

۲-۳- بار ساعتی فیدرهای فشار متوسط

مجموعه داده بار ساعتی فیدرهای فشار متوسط شامل اطلاعات بار ساعتی (برحسب مگاوات) بیش از ۴۰۰ فیدر است و به عنوان داده مکمل استفاده شد تا تأثیر بار بر خاموشی‌ها بررسی شود. پیش‌پردازش شامل ترکیب با داده‌های خاموشی‌های برنامه‌ریزی‌نشده، آماده‌سازی مقادیر ساعتی/روزانه و محاسبه بیشینه بار سالانه برای نرمال‌سازی بود.

از این مجموعه، ویژگی‌های کلیدی شامل بار نرمال‌شده در لحظه خاموشی و میانگین بار نرمال‌شده استخراج شدند. بار نرمال‌شده در لحظه خاموشی که برابر نسبت بار لحظه‌ای به بیشینه بار سالانه فیدر است، برای تشخیص خاموشی در شرایط اوج بار و تنش حرارتی تجهیزات استخراج شد. همچنین، میانگین بار نرمال‌شده در بازه‌های ۶، ۳ و ۱۲ ساعت پیش از خاموشی برای ارزیابی فشار لحظه‌ای، پایدار و تجمعی بر شبکه، اثر انباشته مصرف (مانند افزایش دما و فرسودگی) در نظر گرفته شده است و ریسک را حتی در بار لحظه‌ای پایین آشکار می‌کند. برای اجرای پیش‌پردازش نیز، مقادیر نامعتبر یا صفر با میانگین همان فیدر در ماه مربوط پر شدند.

۲-۴- اطلاعات آب‌وهوایی

مجموعه داده اطلاعات آب‌وهوایی به عنوان داده مکمل در

این مقاله، شامل اطلاعات ساعتی دما و بارندگی برای امورهای عملیاتی مورد مطالعه بوده و از مرجع [۱۸] گردآوری شده است. این داده‌ها شرایط محیطی لحظه وقوع خاموشی را بازنمایی و امکان کشف ارتباط مستقیم میان وضعیت جوی و علل خاموشی را فراهم می‌کنند. در مرحله پیش‌پردازش، داده‌ها با مجموعه خاموشی‌ها ادغام شدند. دمای هوا به شکل پیوسته و بارندگی به صورت دودویی (بارانی/غیربارانی) در نظر گرفته شدند.

۲-۵- جمع‌بندی و ویژگی‌های استفاده‌شده

در نتیجه فرایند پیش‌پردازش و استخراج ویژگی‌ها، مجموعه‌ای از ویژگی‌های پیوسته و گسسته حاصل شد که پس از پاک‌سازی، تبدیل و تجمیع داده‌های خام، مبنای آموزش مدل‌های یادگیری ماشین قرار گرفتند. جدول (۲) نام این ویژگی‌ها را به همراه توضیح مختصر از ماهیت و کاربرد آنها ارائه می‌دهد. افزون بر این، دو ویژگی درصد طول هوایی و زمینی فیدرها به منظور نمایش سهم خطوط هوایی و زیرزمینی در فیدر محل خاموشی در نظر گرفته شدند، زیرا ساختار شبکه تأثیری مستقیم بر علل و الگوهای خاموشی دارد؛ به گونه‌ای که بخش‌های هوایی بیشتر در معرض عوامل محیطی و بخش‌های زمینی عمدتاً مستعد عیوب فنی کابل هستند.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (۳)$$

همچنین، برای آموزش مدل، از تابع هزینه لگاریتمی یا Binary Cross-Entropy استفاده می‌کنیم. معمولاً، پس از آموزش، برای تعیین کلاس نمونه جدید، مطابق رابطه ۴، از آستانه ۰/۵ استفاده می‌شود که البته آستانه ممکن است بسته به مسأله تغییر کند تا حساسیت و دقت مدل بهینه شود.

$$\hat{y} = \begin{cases} 1 & \text{if } P(y=1 | \mathbf{x}) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (۴)$$

مزایای LR شامل سادگی، سرعت محاسباتی زیاد، تفسیرپذیری عالی (از طریق ضرایب نشان‌دهنده جهت و شدت تأثیر ویژگی‌ها بر احتمال کلاس مثبت w در معادله ۱ و ۲)، مقاومت متوسط در برابر بیش برآزش و عملکرد مناسب در داده‌های حجیم است [۱۹].

۳-۲- مدل‌های گروهی

مدل‌های گروهی^{۱۱} با ترکیب چندین مدل پایه، دقت، پایداری و تعمیم‌پذیری مدل‌های یادگیری ماشین را افزایش می‌دهند و با کاهش سوگیری نسبت به کلاس‌های خاص و خطاهای تصادفی، عملکردی بهتر نسبت به مدل‌های منفرد ارائه می‌کنند [۲۰] و [۲۱]. این رویکردها به ویژه در مسائل طبقه‌بندی پیچیده؛ مانند شناسایی علت خاموشی‌ها، به دلیل توانایی در مدیریت داده‌های نامتعادل و نویزی، مؤثر هستند. در این مقاله، مدل‌های گروهی مبتنی بر LR به عنوان مدل‌های پایه برای مسائل طبقه‌بندی دوتایی (یک علت در برابر سایر علل)^{۱۲} به کار گرفته می‌شوند؛ زیرا سادگی، تفسیرپذیری و سرعت LR با توان ترکیب در چارچوب‌های یادگیری گروهی همخوانی دارد و دقت شناسایی علل خاموشی در لحظه وقوع را بهبود می‌بخشد [۲۰] و [۲۱]. به طور مشخص، در رویکرد Stacking، چندین مدل پایه آموزش داده می‌شوند و خروجی آنها به عنوان ورودی یک مدل ترکیبی^{۱۳} برای تولید پیش‌بینی نهایی استفاده می‌شود [۲۰]. این چارچوب در مقایسه با مدل‌های منفرد LR، مقاومت بیشتری در برابر بیش برآزش دارد و امکان اولویت‌بندی علل محتمل خاموشی را نیز فراهم می‌کند.

جدول (۲): ویژگی‌های استفاده‌شده برای آموزش مدل

نام خصیصه	توضیحات
منطقه عملیاتی	گسسته، ۲۴ منطقه عملیاتی
فصل	گسسته، نگاهت ماه به فصل
بازه زمانی وقوع خاموشی	گسسته، تقسیم روز به ۴ دوره زمانی
نقطه قطع	گسسته (قطع یا عدم از ابتدای فیدر)
آخرین خاموشی گذرا در فیدر	پیوسته، فاصله زمانی آخرین خاموشی گذرا از خاموشی رخ داده در همان فیدر
سابقه تعمیرات تجهیز X در فیدر	پیوسته، فاصله زمانی آخرین تعمیر و نگهداری در تجهیز X در همان فیدر
تعداد خاموشی‌های گذرا در ۳۰ روز گذشته در فیدر	تعداد خاموشی‌های گذرای رخ داده در همان فیدر در ۳۰ روز قبل از وقوع خاموشی
بار لحظه‌ای و میانگین نرمال‌شده	پیوسته، بار لحظه‌ای و میانگین نرمال‌شده ۶، ۳ و ۱۲ ساعته (تقسیم بار ابتدای فیدر به پیک بار سالانه فیدر)
درصد طول هوایی و زمینی فیدر	پیوسته، درصد طول فیدر فشار متوسط که به صورت خطوط هوایی یا کابل‌های زیرزمینی قرار دارد.
دمای هوا و بارندگی	دمای هوا، پیوسته است و بارندگی/عدم بارندگی، خصیصه‌ای دودویی

۳- مدل‌های طبقه‌بندی

در این بخش، مبانی نظری مرتبط با مدل‌های طبقه‌بندی مبتنی بر رگرسیون لجستیک و مدل‌های گروهی به اختصار ارائه می‌شود.

۳-۱- رگرسیون لجستیک

رگرسیون لجستیک (LR) یکی از الگوریتم‌های پرکاربرد در مسائل طبقه‌بندی دوتایی است که هدف آن پیش‌بینی احتمال تعلق یک نمونه به یکی از دو کلاس (طبقه‌بندی دوتایی)^{۱۴} است. خروجی LR به بازه $[0, 1]$ محدود می‌شود تا به عنوان احتمال تفسیر شود.

احتمال تعلق نمونه به کلاس مثبت (یعنی $y=1$) به صورت معادله ۱ محاسبه می‌شود و احتمال کلاس منفی نیز با معادله ۲ مشخص می‌شود. در آنها، $\omega^T \mathbf{x} + b$ ، با $\mathbf{x} = [x_1, x_2, \dots, x_n]$ به عنوان بردار ویژگی‌ها، $\omega = [\omega_1, \omega_2, \dots, \omega_n]$ به عنوان بردار ضرایب (وزن‌ها) و b به عنوان بایاس شناخته می‌شود. σ نیز نشان‌دهنده تابع سیگموید است که در معادله ۳ نشان داده شده است.

$$P(y=1 | \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b) \quad (۱)$$

$$P(y=0 | \mathbf{x}) = 1 - \sigma(\mathbf{w}^T \mathbf{x} + b) \quad (۲)$$

است که در آن، K تعداد ضرایب نمونه برداری است.

۴- روش پیشنهادی

در این بخش، روش پیشنهادی برای طبقه بندی علل خاموشی در سیستم توزیع ارائه می شود. در این روش، برای هر یک از علل خاموشی، یک مدل گروهی مجزا آموزش داده می شود که به صورت دودویی پیش بینی می کند آیا آن علت خاص رخ داده است یا خیر. در نهایت، با ترکیب خروجی همه مدل های گروهی، محتمل ترین علت (یا علل) خاموشی تعیین و بر اساس احتمال نزولی اولویت بندی و ارائه می شود.

۴-۱- آموزش مدل های گروهی برای هر یک از علل خاموشی

در این روش، برای هر علت خاموشی، یک مجموعه داده فرعی ایجاد می شود که رکوردهای آن علت به عنوان کلاس مثبت و سایرین به عنوان کلاس منفی برچسب گذاری می شوند. به این ترتیب، چندین مدل دودویی تخصصی (به جای یک مدل چندکلاسه واحد) آموزش داده می شوند که هر کدام بر تشخیص یک علت خاص تمرکز دارد.

برای مدیریت نامتعادلی، ضرایب نمونه برداری (α_i) (در این مقاله، $0/5$ ، $0/7$ ، 1 ، $1/3$ و $1/5$) متنوعی تعریف می شود تا نسبت اکثریت به اقلیت تغییر کند؛ برای هر ضریب و در هر یک از مجموعه داده های فرعی بیان شده، یک زیرمدل LR آموزش دیده می شود و سپس، زیرمدل ها در یک مدل گروهی با رأی گیری نرم^{۱۴} ترکیب می شوند. خروجی نهایی میانگین احتمالات زیرمدل هاست که الگوهای متمایز هر علت را مؤثرتر می آموزد. مدل گروهی می تواند برای تشخیص آنلاین وقوع/عدم وقوع آن علت به کار رود. خروجی نهایی مدل گروهی برای نمونه x برای علت خاموشی c به صورت معادله ۵ محاسبه می شود.

$$P_{\text{ensemble}}^c(y=1|x) = \sum_{k=1}^K \frac{1}{K} P_k^c(y=1|x; \alpha_k) \quad (5)$$

در رابطه ۵، $P_k^c(y=1|x; \alpha_k)$ احتمال پیش بینی شده توسط زیرمدل k با ضریب نمونه برداری α_k است و در اینجا فرض می شود وزن تمامی مدل های به دست آمده از ضرایب مختلف نمونه برداری برای هر علت خاموشی (c) برابر $\frac{1}{K}$

۴-۲- تعیین آستانه بهینه برای تشخیص وقوع علت خاموشی

خروجی مدل های LR احتمالی بین ۰ تا ۱ است که برای تبدیل به پیش بینی قطعی، از یک آستانه استفاده می شود. آستانه پیش فرض ساده $0/5$ است و فرض تعادل کلاس ها و هزینه برابر خطاها را دارد، اما در داده های نامتعادل علت خاموشی، این آستانه بهینه نیست و TPR یا TNR را کاهش می دهد. برای برطرف کردن این مشکل، از روش شاخص یودن^{۱۵} برای تعیین آستانه بهینه هر مدل استفاده شد که طی آن، آستانه ای که شاخص یودن را در مجموعه اعتبارسنجی بیشینه کند، به عنوان آستانه بهینه انتخاب می شود [۲۲]. شاخص یودن در رابطه ۶ نشان داده شده است.

$$J = \text{TPR} + \text{TNR} - 1 \quad (6)$$

روند تشخیص آستانه بهینه نیز به صورت آزمون و خطا خواهد بود که طی آن، آستانه های مختلف از ۰ تا ۱ آزموده می شوند و در نهایت، آستانه بهینه انتخاب می شود. آستانه بهینه مدل گروهی کلاس c برای رکورد i ام به صورت زیر استفاده می شود که در آن، θ_c^* نشانگر آستانه بهینه برای مدل گروهی کلاس c است.

$$\hat{y}_i = \begin{cases} 1 & \text{if } P_{\text{ensemble}}^c(y=1|x_i) \geq \theta_c^* \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

هر مدل گروهی ساخته شده می تواند مستقلاً برای تشخیص دودویی یک علت خاص استفاده شود، اما در کاربرد اصلی سیستم، تمام مدل ها هم زمان اجرا، احتمالات محاسبه و علل بر اساس احتمال نزولی اولویت بندی می شوند. علل برتر (برای مثال، ۳ اولویت بالاتر از همه) به عنوان محتمل ترین پیشنهاد می شود تا بررسی میدانی از آنها آغاز شود.

۴-۳- طبقه بندی نهایی با ترکیب خروجی مدل های گروهی

در روش پیشنهادی، خروجی های مدل های گروهی

دودویی (معادله ۵) به صورت موازی برای تصمیم‌گیری نهایی استفاده می‌شوند. احتمالات پیش‌بینی شده با وزن‌دهی بر اساس AUC هر مدل روی مجموعه اعتبارسنجی اصلاح می‌شوند تا مدل‌های دقیق‌تر تأثیر بیشتری داشته باشند. فرمول امتیاز وزنی برای هر علت c به صورت معادله (۸) محاسبه می‌شود. $AUC_c^{validation}$ مقدار AUC به‌دست‌آمده از مجموعه اعتبارسنجی برای مدل دودویی مربوط به کلاس c است. سپس، علل بر اساس امتیاز وزنی نزولی مرتب می‌شوند و محتمل‌ترین علت (Top-1) همراه با فهرست چند علت برتر (Top-N) استخراج می‌شود.

$$Score_c(x_i) = P_{ensemble}^c(y=1|x_i) \times AUC_c^{validation} \quad (8)$$

رویکرد پیشنهادی مزایای کلیدی زیر را داراست:

- ۱) وزن‌دهی بر اساس AUC که کیفیت نسبی مدل‌ها را در نظر می‌گیرد و دقت پیش‌بینی را بهبود می‌بخشد؛
- ۲) ارائه Top-N به جای یک علت قطعی انعطاف‌پذیری بیشتری در تصمیم‌گیری عملی ایجاد می‌کند؛
- ۳) کمک به بهره‌برداران برای اولویت‌بندی بررسی‌های میدانی و مدیریت منابع با ارائه دید کلی از سناریوهای محتمل در لحظه وقوع خاموشی.

شبه‌کد شماره ۱ نشانگر چگونگی استفاده از روش ارائه‌شده در این بخش برای طبقه‌بندی علت خاموشی است.

شبه‌کد شماره ۱: آموزش و آزمایش روش پیشنهادی مبتنی بر مدل‌های گروهی برای شناسایی علت خاموشی در سیستم توزیع

ورودی: مجموعه داده‌های خاموشی، تعمیرات، بار فیدرها، داده‌های آب‌وهوایی (خصیصه‌های استخراج‌شده و برچسب‌ها)، فهرست علل خاموشی
برای هر علت خاموشی در فهرست علل خاموشی:
۱. ساخت مجموعه داده فرعی:
رکوردهای علت خاموشی مدنظر ← کلاس اصلی

سایر رکوردها ← کلاس دیگران

۲. برای هر ضریب نمونه‌برداری α :

اعمال نمونه‌برداری تصادفی با نسبت α

آموزش یک زیرمدل دوتایی روی مجموعه داده آموزش ساخته‌شده

۳. ترکیب تمام زیرمدل‌ها با رأی‌گیری نرم ← ایجاد مدل گروهی برای علت
۴. تقسیم داده‌ها به مجموعه آموزش و اعتبارسنجی
۵. پیش‌بینی احتمال وقوع کلاس هدف روی مجموعه اعتبارسنجی

۶. محاسبه منحنی ROC و شاخص یودن
۷. انتخاب آستانه بهینه که بالاترین شاخص یودن را دارد
۸. ذخیره مدل گروهی و آستانه بهینه

برای هر رکورد جدید:

۱. برای هر مدل گروهی (هر علت):
پیش‌بینی احتمال وقوع علت و استفاده از آستانه بهینه برای تعیین وقوع یا عدم وقوع خاموشی با آن علت به‌خصوص
ضرب احتمال به‌دست‌آمده از هر مدل در وزن مدل (AUC اعتبارسنجی)
۲. مرتب‌سازی علل بر اساس مقادیر وزنی
۳. خروجی:
Top-1: محتمل‌ترین علت خاموشی
Top-N: فهرست چند علت محتمل دیگر
خروجی: فهرست علل احتمالی هر رکورد، به همراه Top-1 و Top-N و نیز، علل اعلام‌شده خاموشی با استفاده از آستانه‌های بهینه

۵- نتایج و بحث در آنها

در این بخش، ابتدا نتایج حاصل از آموزش و ترکیب مدل‌های گروهی مبتنی بر LR و تعیین آستانه تصمیم‌گیری بهینه ارائه و با روش‌های مبتنی بر درخت تصمیم [۱۵] مقایسه می‌شود. سپس، عملکرد روش پیشنهادی در ترکیب و وزن‌دهی خروجی مدل‌ها و دقت آن در پیش‌بینی اولویت علل خاموشی ارزیابی و با روش [۱۴] مقایسه می‌شود. در نهایت، روند پیشنهادی اولویت‌دهی علل خاموشی‌های گذرا بررسی و نتایج آن با روش مبتنی بر قواعد انجمنی ارائه‌شده در [۱۶] مقایسه

ایجاد کرد (جدول ۳). نتایج نشان داد این بهینه‌سازی، TPR کلاس‌های اقلیت را بدون کاهش چشمگیر TNR افزایش داد. مقایسه با [۱۵] که از مدل‌های گروهی مبتنی بر درخت تصمیم C4.5 استفاده کرده است، برتری روش پیشنهادی را نشان می‌دهد. برای عللی مانند کابل، سرکابل، جمپر و خط توزیع هوایی، مدل پیشنهادی TPR بالاتری به دست آورده است. همچنین، برای کلاس‌هایی مانند حیوانات، شرایط بد آب‌وهوایی و پوشش گیاهی که در [۱۵] بررسی نشده‌اند، مدل پیشنهادی به AUC بالاتر از ۰/۸۲ دست یافته است. این بهبود عمدتاً ناشی از استفاده از LR به جای C4.5 با مقاومت بهتر در برابر نامتعادلی داده‌ها، به‌کارگیری ویژگی‌های غنی‌تر مانند فاصله زمانی تا تعمیرات اخیر و میانگین بار ساعات پیشین و نیز استفاده از رأی‌گیری نرم در ترکیب مدل‌های گروهی است؛ در حالی که [۱۵] از رأی‌گیری سخت^{۲۰} بهره برده است.

از دیدگاه کاربردی، روش پیشنهادی، با وجود استفاده از چندین مدل گروهی، همچنان بسیار سبک و کارآمد است. زمان کل آموزش تمام مدل‌های گروهی برای ۱۶ کلاس اصلی روی داده‌های واقعی (بیش از ۵۰ هزار رکورد که ۸۰ درصد آن برای آموزش و اعتبارسنجی استفاده شده است) حدود ۶۴/۵ ثانیه (معادل ۱/۰۷ دقیقه) بوده است؛ یعنی میانگین زمان آموزش برای هر کلاس حدود ۴ ثانیه است؛ هرچند بخشی قابل ملاحظه از زمان آموزش به کلاس بزرگ و پرتکرار «عملکرد یا اشکال در کات‌اوت فیوز» اختصاص یافته است. زمان استنتاج^{۲۱} نیز بسیار پایین است؛ میانگین زمان پردازش برای ۱۰۰ نمونه حدود ۰/۲۸۴ ثانیه بوده است که معادل حدود ۲/۸۴ میلی‌ثانیه به‌ازای هر نمونه می‌شود. این سرعت زیاد امکان استفاده کاملاً آنلاین روش را در لحظه وقوع خاموشی فراهم می‌کند؛ به گونه‌ای که بلافاصله پس از ثبت اطلاعات خاموشی، سیستم می‌تواند علل محتمل را اولویت‌بندی کند و پیشنهادهای عملی را به بهره‌بردار ارائه دهد. با توجه به سادگی ذاتی رگرسیون لجستیک و عدم نیاز به منابع محاسباتی سنگین، روش پیشنهادی به راحتی قابلیت اجرا در محیط‌های واقعی شبکه‌های توزیع نیروی برق را دارد و می‌تواند بدون نیاز به سخت‌افزارهای گران‌قیمت یا سرورهای قدرتمند، در مراکز کنترل و دیسپاچینگ استفاده شود.

برای بررسی دقیق‌تر تأثیر رویکرد یادگیری گروهی، یک

می‌شود. ضمناً، برای اجرای تمام آزمایش‌ها و ارزیابی عملکرد روش پیشنهادی، از یک سیستم با مشخصات زیر استفاده شد: پردازنده Intel Core i7-8700K با فرکانس ۳/۷ گیگاهرتز، ۲۴ گیگابایت رم و محیط اجرای پایتون نسخه ۳.۹.۱۲. تمامی زمان‌های گزارش‌شده در این بخش بر اساس اجرای کد روی این سیستم اندازه‌گیری شده‌اند.

۱-۵- طبقه‌بندی علل خاموشی با استفاده از مدل‌های گروهی مبتنی بر LR

در این بخش، فرایند آموزش مدل‌های اولیه تشریح می‌شود و با بهره‌گیری از مفهوم مدل‌های گروهی، مدل‌های LR آموزش‌دیده برای هر یک از علل خاموشی با یکدیگر ترکیب می‌شوند. سپس، آستانه تصمیم‌گیری مدل با استفاده از معیار یودن به صورت بهینه تعیین و عملکرد مدل‌های گروهی پیشنهادی مبتنی بر LR با روش ارائه‌شده در مرجع [۱۵] مقایسه می‌شود که مبتنی بر روشی بر پایه درخت تصمیم است.

ابتدا، سه برجسب غیراستاندارد («نامربوط به توزیع»، «عیب نامشخص» و «حاوی اثر انسانی»)، به علت نامرتب بودن به توزیع یا عدم اثرگذاری عوامل غیرقطعی روی وقوع آنها، ادغام و از آموزش حذف شدند. پیش‌پردازش شامل نرمال‌سازی Z-score برای ویژگی‌های پیوسته و OneHotEncoder برای ویژگی‌های گسسته اجرا شد. داده‌ها به نسبت ۸۰ درصد آموزش و ۲۰ درصد آزمایش تقسیم شدند. سپس، روی داده‌های آموزش، تشخیص ناهنجاری^{۱۶} با جنگل انزوا^{۱۷} اعمال شد. پس از پالایش، داده‌های آموزش مجدداً به ۸۰ درصد آموزش نهایی و ۲۰ درصد اعتبارسنجی تقسیم شدند. مجموعه داده‌ها با نسبت‌های مختلف (α_i) اکثریت/اقلیت ساخته شد. برای هر زیرمجموعه، انتخاب فرآپارامتر^{۱۸} های مدل LR با جست‌وجوی شبکه‌ای^{۱۹} اجرا شد. در ابتدا، با آستانه پیش‌فرض ۰/۵، مدل‌ها عملکردی مناسب در کلاس‌های اکثریت در زمان آزمایش نشان دادند، اما در کلاس‌های اقلیت (برای مثال، پوشش گیاهی با ۳۵ رکورد) TPR پایین بود. برای بهبود این امر، آستانه بهینه با روش شاخص یودن تعیین شد که تعادل بهتری بین TPR و TNR

± 2 انحراف معیار (تقریباً معادل بازه اطمینان ۹۵ درصد) برای معیارهای TPR، TNR و AUC در جدول (۴) ارائه شده است. همان‌طور که مشاهده می‌شود، روش پیشنهادی در کلاس‌های پرتکرار (مانند عملکرد یا اشکال در کات‌اوت فیوز و خاموشی گذرا) دارای پایداری بسیار زیادی است. با این حال، در کلاس‌های اقلیت انحراف معیار بیشتری مشاهده می‌شود که عمدتاً ناشی از حجم بسیار محدود نمونه‌های این کلاس‌ها است. با وجود این نوسان، رویکرد یادگیری گروهی با نمونه‌برداری مجدد هدفمند توانسته است الگوهای پنهان کلاس‌های اقلیت را بهتر شناسایی کند و بهبودهایی قبل از ملاحظه در AUC و تعادل بین TPR و TNR ایجاد کند.

۲-۵- اولویت‌دهی به علل احتمالی خاموشی‌ها

در این بخش، عملکرد روش پیشنهادی بر اساس پیش‌بینی علت با اولویت اول و همچنین، قرارگیری علت واقعی در میان سه و پنج اولویت نخست ارزیابی می‌شود. این اولویت‌ها با استفاده از چارچوب مدل‌های گروهی و بر مبنای خروجی‌های احتمالی وزن‌دهی‌شده بر اساس AUC هر مدل به دست آمده‌اند. نتایج این ارزیابی در جدول (۵) ارائه شده است و نشان می‌دهد علت واقعی در چه درصدی از نمونه‌های مجموعه آزمایش با اولویت اول منطبق است یا در میان سه و پنج اولویت نخست قرار گرفته است.

مدل پایه رگرسیون لجستیک تک نیز با راهبرد One-vs-Rest (اما بدون نمونه‌برداری مجدد و ترکیب گروهی) روی همان مجموعه داده و ویژگی‌ها آموزش داده شد. مقایسه نتایج نشان می‌دهد مدل گروهی در کلاس‌های اکثریت (مانند کات‌اوت فیوز و خاموشی گذرا) عملکردی مشابه یا کم‌بهتری داشته است. با این حال، مهم‌ترین دستاورد رویکرد پیشنهادی در کلاس‌های اقلیت مشاهده شد. مدل گروهی توانست در بیشتر کلاس‌های کم‌تکرار AUC را تا حدی مناسب بهبود دهد. برای مثال، در کلاس پایه AUC از ۰/۷۶۳ به ۰/۷۹۹، در کلاس حیوانات از ۰/۸۱۷ به ۰/۸۲۵، و در کلاس پوشش گیاهی از ۰/۸۷۳ به ۰/۸۸۳ افزایش یافت. فراتر از افزایش AUC، رویکرد گروهی تعادلی بهتر بین TPR و TNR برقرار کرد. در کلاس‌هایی مانند شرایط بد آب‌وهوایی، حیوانات و مشکلات سمت مشترکین، مدل گروهی هم‌زمان TPR و TNR را بهبود بخشید یا تعادلی منطقی‌تر بین آنها ایجاد کرد. این بهبود تعادل، همراه با افزایش پایداری^{۲۲} نتایج در برابر تغییرات توزیع داده، نشان‌دهنده برتری مدل گروهی نسبت به مدل تک LR است. برای ارزیابی پایداری و قابلیت اعتماد روش پیشنهادی نیز، مدل گروهی برای هر علت خاموشی با تغییر ۱۰۰ بار seed تصادفی برای تقسیم‌بندی داده‌ها به مجموعه‌های آموزش و آزمون اجرا و ارزیابی شد. نتایج به صورت میانگین

جدول ۳- عملکرد مدل‌های گروهی مربوط به هر علت خاموشی در مجموعه‌ی آزمایش و مقایسه عملکرد آن‌ها با روش پیشنهادی [۱۵]

علت خاموشی	تعداد رکوردها	استانده بهینه	TPR [۱۵]	TNR	TNR [۱۵]	AUC	AUC [۱۵]
عملکرد یا اشکال در کات اوت فیوز	۷۴۱۰	۰/۵۰۹	۰/۹۸	۰/۹۵۵	۰/۷۴۴	۰/۷۳۳	۰/۸۶۱
خاموشی گذرا	۱۱۸۳	۰/۶۵	۰/۹۸۵	۰/۹۱۹	۰/۹۲۹	۰/۹۲۴	۰/۹۲۳
حیوانات	۲۴۴	۰/۵۶۹	۰/۷۷۹	۰/۶۸	۰/۷۷۱	۰/۶۹۵	۰/۶۸۳
شرایط بد آب‌وهوایی	۱۳۸	۰/۳۹۹	۰/۸۴۱	۰/۸۱۹	۰/۸۳۵	۰/۸۶۳	۰/۹۰۵
چمبر	۱۳۷	۰/۴۹۷	۰/۶۷۹	۰/۵۹۹	۰/۶۶۷	۰/۵۸۲	۰/۷۱۷
سرکابل	۱۳۲	۰/۴۴۶	۰/۹۳۲	۰/۸۳۳	۰/۸۵۱	۰/۸۲۸	۰/۹۳۷
تجهیزات پست توزیع	۱۰۹	۰/۵۵۸	۰/۶۸۸	۰/۶۳۳	۰/۷۶	۰/۶۳	۰/۸۰۹
ادوات کلیدزنی (بریکر و سکسیونر)	۱۰۴	۰/۴۵۸	۰/۷۵	۰/۶۰۶	۰/۶۶	۰/۶۲	۰/۷۵۴
کابل	۱۰۴	۰/۳۴۵	۰/۹۰۴	۰/۷۹۸	۰/۸۵۷	۰/۸۵	۰/۹۲۷
مشکلات سمت مشترکین	۷۶	۰/۴۰۶	۰/۷۷۶	۰/۶۹۷	۰/۷۱	۰/۷۰۶	۰/۸۲۴
ترانسفورماتور	۷۲	۰/۶۲۵	۰/۵۶۹	۰/۶۱۱	۰/۸۴۸	۰/۶۱۷	۰/۷۴۸
خط توزیع هوایی و متعلقات آن	۵۹	۰/۴۷	۰/۶۴۴	۰/۷۱۲	۰/۷۵۸	۰/۶۳۹	۰/۷۳
پایه	۴۸	۰/۳۴	۰/۷۷۱	۰/۶۸۸	۰/۷۰۷	۰/۶۵۶	۰/۷۹۹
مقره	۴۰	۰/۴۴۶	۰/۸۲۵	۰/۶۲۵	۰/۷۸۴	۰/۷۱۷	۰/۸۶۱
پوشش گیاهی	۳۵	۰/۴۱۱	۰/۸۵۷	۰/۷۴۳	۰/۷۸۴	۰/۷۹۳	۰/۸۸۳
برق‌گیر	۳۴	۰/۳۹۶	۰/۷۰۶	۰/۶۱۸	۰/۶۳۵	۰/۶۰۳	۰/۷۴۹

جدول (۴): حدود معیارهای ارزیابی مدل‌های گروهی در اجرای ۱۰۰ باره فرایند آموزش و آزمایش (بازه اطمینان ۹۵ درصد)

کلاس	تعداد رکوردهای آزمایش	حدود TPR	حدود TNR	حدود AUC
عملکرد یا اشکال در کات اوت فیوز	۷۴۱۰	0.984 ± 0.014	0.741 ± 0.015	0.920 ± 0.005
خاموشی گذرا	۱۱۸۳	0.985 ± 0.008	0.927 ± 0.004	0.973 ± 0.003
حیوانات	۲۴۴	0.825 ± 0.077	0.744 ± 0.054	0.832 ± 0.023
شرایط بد آب‌وهوایی	۱۳۸	0.791 ± 0.242	0.793 ± 0.206	0.860 ± 0.205
جمپر	۱۳۷	0.798 ± 0.273	0.711 ± 0.246	0.768 ± 0.206
سرکابل	۱۳۲	0.913 ± 0.080	0.833 ± 0.050	0.931 ± 0.016
تجهیزات پست توزیع	۱۰۹	0.731 ± 0.193	0.797 ± 0.158	0.792 ± 0.040
ادوات کلیدزنی (بریکر و سکسیونر)	۱۰۴	0.759 ± 0.18	0.774 ± 0.159	0.724 ± 0.052
کابل	۱۰۴	0.914 ± 0.081	0.859 ± 0.055	0.946 ± 0.020
مشکلات سمت مشترکین	۷۶	0.763 ± 0.166	0.729 ± 0.138	0.827 ± 0.038
ترانسفورماتور	۷۲	0.761 ± 0.270	0.780 ± 0.243	0.736 ± 0.051
خط توزیع هوایی و متعلقات آن	۵۹	0.701 ± 0.198	0.732 ± 0.167	0.783 ± 0.048
پایه	۴۸	0.728 ± 0.204	0.759 ± 0.118	0.813 ± 0.057
مقره	۴۰	0.751 ± 0.227	0.756 ± 0.151	0.832 ± 0.050
پوشش گیاهی	۳۵	0.798 ± 0.178	0.780 ± 0.109	0.853 ± 0.057
برق‌گیر	۳۴	0.736 ± 0.317	0.763 ± 0.255	0.715 ± 0.083

جدول (۵): عملکرد روش پیشنهادی برای شناسایی علت خاموشی و دقت حضور در اولویت‌های اول در مجموعه آزمایش و مقایسه

عملکرد آنها با روش پیشنهادی [۱۴]

علت خاموشی	تعداد رکوردهای مثبت	دقت شناسایی به عنوان اولویت اول (%)	دقت حضور در ۳ اولویت اول (%)	دقت حضور در ۵ اولویت اول (%)	Recall (%) [۱۴]
کات‌اوت فیوز	۷۴۱۰	۷۶/۴۵	۹۴/۴	۹۷/۶۷	۹۶/۸
خاموشی گذرا	۱۱۸۳	۷۷/۳۵	۹۳/۶۶	۹۷/۸	۲۱/۸
حیوانات	۲۴۴	۵۲/۴۶	۸۰/۷۴	۸۴/۰۲	۱۳/۵
شرایط بد آب‌وهوایی	۱۳۸	۳۲/۶۱	۶۷/۳۹	۷۳/۹۱	۴۶/۴
جمپر	۱۳۷	۲/۱۹	۲۱/۱۷	۴۸/۹۱	۰/۷
سرکابل	۱۳۲	۱۹/۷	۶۵/۱۵	۸۰/۳	۱۱/۴
پست توزیع	۱۰۹	۲۱/۱	۷۱/۵۶	۸۱/۶۵	۰
کابل	۱۰۴	۵۱/۹۲	۷۶/۹۲	۸۲/۶۹	۱۹/۲
ادوات کلیدزنی	۱۰۴	۰	۲۰/۱۹	۴۶/۱۵	۰
مشترکین	۷۶	۰	۲۵	۶۳/۱۶	۰
ترانسفورماتور	۷۲	۴/۱۷	۵۰	۷۹/۱۷	۰
خط توزیع هوایی	۵۹	۰	۸/۴۷	۳۳/۹	۰
پایه	۴۸	۰	۱۲/۵	۳۹/۵۸	۰
مقره	۴۰	۰	۲۲/۵	۴۲/۵	۰
پوشش گیاهی	۳۵	۵/۷۱	۴۲/۸۶	۵۷/۱۴	۲/۹
برق‌گیر	۳۴	۰	۲۰/۵۹	۵۰	۰

همچنین، روش پیشنهادی اولویت‌دهی به علل خاموشی با روش [۱۴] نیز مقایسه می‌شود که بر استفاده از ماتریس هزینه^{۳۳} تمرکز دارد. [۱۴] بدون نمونه‌گیری مجدد، عدم تعادل کلاس‌ها را تا حدی مدیریت و خطاهای با پیامدهای مهم‌تر، مانند تشخیص نادرست خاموشی‌های ناشی از حیوانات،

پوشش گیاهی یا شرایط جوی، را شدیدتر جریمه می‌کند. همچنین، با ترکیب خروجی درخت تصمیم و جنگل تصادفی در یک مدل گروهی، بدون آنکه متعادل‌سازی صریح داده‌ها انجام شود، سعی در بهبود عملکرد طبقه‌بندی داشته است. مقایسه روش پیشنهادی با رویکرد [۱۴] نشان‌دهنده برتری

همچنین، روش پیشنهادی اولویت‌دهی به علل خاموشی با روش [۱۴] نیز مقایسه می‌شود که بر استفاده از ماتریس هزینه^{۳۳} تمرکز دارد. [۱۴] بدون نمونه‌گیری مجدد، عدم تعادل کلاس‌ها را تا حدی مدیریت و خطاهای با پیامدهای مهم‌تر، مانند تشخیص نادرست خاموشی‌های ناشی از حیوانات،

قابل ملاحظه آن، به ویژه در کلاس‌های اقلیت و نامتعادل است. این بهبود عمدتاً به دلیل استفاده از مدل‌های گروهی مبتنی بر LR با وزن‌دهی بر اساس AUC، استخراج ویژگی‌های غنی‌تر و مهم‌تر از همه، مقابله با نامتعادلی شدید داده از طریق استفاده از رهیافت One-vs-Rest و مدل‌های گروهی است. در حالی که روش [۱۴] بر جنگل تصادفی و درخت تصمیم تکیه دارد و در کلاس‌های نادر عملکردی ضعیف‌تر دارد، روش پیشنهادی نه فقط دقت حضور در اولویت اول را افزایش می‌دهد، بلکه با اولویت‌بندی Top-3 و Top-5 (بالای ۷۰ درصد در بسیاری از کلاس‌ها)، کاربرد عملی بیشتری برای بهره‌بردار فراهم می‌آورد. اجرای ۱۰۰ باره فرایند آموزش و آزمایش نیز نشان داد عموماً کلاس‌های پرجمعیت به علت تعداد زیاد رکوردشان در دقت تشخیص علت خاموشی به عنوان اولویت اول پایداری بسیار زیادی دارند (برای مثال، بازه اطمینان ۹۵ درصد معادل $0.801 \pm$ برای دقت تشخیص اولویت اول برای کلاس خاموشی‌های گذرا). البته، مشابه جدول (۴)، دقت تشخیص علت اقلیت مانند پوشش گیاهی، تنوع بیشتری دارد (برای مثال، بازه اطمینان ۹۵ درصد معادل $0.586 \pm$ برای کلاس پوشش گیاهی برای دقت حضور در ۵ اولویت اول). این تنوع در دقت شناسایی این علت خاموشی به ویژه ناشی از تعداد بسیار کم آن در مجموعه داده اصلی است.

۳-۵- شناسایی علت احتمالی خاموشی‌های گذرا در سیستم توزیع

خاموشی‌های گذرا معمولاً بدون ثبت علت واقعی (فقط به عنوان «عیب گذرا») باقی می‌مانند، اما تکرار آنها پیش‌نشانگر خاموشی‌های دائم است. شناسایی علت واقعی این خاموشی‌ها امکان اقدامات پیشگیرانه هدفمند را فراهم می‌کند تا احتمال تبدیل به خاموشی‌های پایدار کاهش یابد. در این بخش، مدل اولویت‌دهی علت خاموشی بر داده‌های خاموشی‌های گذرا اعمال می‌شود و علت محتمل اولویت‌بندی می‌شوند. نتایج نیز با روش [۱۶] مقایسه می‌شود که مبتنی بر الگوریتم Apriori برای استخراج قواعد انجمنی و شباهت‌سنجی خاموشی‌های دائم و گذراست. برای ارزیابی منصفانه با معیار یکسان، از کاهش احتمالی ENS، ناشی از علت حدس

زده‌شده در ۶۰ روز پس از خاموشی گذرا در همان فیدر استفاده شد. این همان معیاری است که در [۱۶] نیز به کار رفته است. ایده اصلی بر شباهت شرایط حاکم بر دو نوع خاموشی استوار است؛ رویکرد پیشنهادی سه علت برتر را برای هر خاموشی گذرا کشف می‌کند و در هدف نهایی مشابه [۱۶] است. با وجود این، روش پیشنهادی از مدل‌های گروهی به جای قواعد انجمنی [۱۶] بهره می‌برد.

نتایج مقایسه روش پیشنهادی در اولویت‌دهی به علل احتمالی خاموشی‌های گذرا (۵۹۱۵ خاموشی گذرا در فاصله سال‌های ۱۳۹۹ تا ۱۴۰۳)، ENS حادث‌شده ناشی از خاموشی دائم در ۶۰ روز بعد از وقوع خاموشی گذرا به همان علت شناسایی‌شده در همان فیدر در روش پیشنهادی و روش [۱۶] در جدول (۶) آورده شده است. مجموع ENS سال‌های ۱۳۹۹ تا ۱۴۰۳ مقدار $7638/2$ مگاوات ساعت است. مشخصاً، بهره‌بردار با استفاده از علل شناسایی‌شده برای خاموشی‌های گذرا می‌توانسته از وقوع $1477/2$ مگاوات ساعت ($19/34$ درصد ENS کل) و در صورت استفاده از روش [۱۶] می‌توانسته از وقوع 1166 مگاوات ساعت ($15/27$ درصد ENS کل) جلوگیری کند؛ زیرا احتمالاً این مقدار پس از گذشت زمانی کوتاه در همان فیدر به خاموشی دائم از همان نوع منجر می‌شده است. همچنین، جدول (۶) نشان می‌دهد ترکیب اولویت‌های روش پیشنهادی از نظر فنی و عملی کاملاً منطقی است. حضور عللی مانند مقره، پایه، جمپر، کلیدزنی و سمت مشترکین در سه اولویت اول با واقعیت شبکه توزیع و تجربه بهره‌برداران هم‌خوانی دارد. توزیع نسبتاً یکنواخت علل در سه اولویت برتر، عدم قطعیت طبیعی شناسایی علت خاموشی گذرا را منعکس کرده و مهر تأییدی بر واقع‌بینانه بودن خروجی روش پیشنهادی است.

برتری روش پیشنهادی نسبت به [۱۶] به تفاوت منطق مدل‌سازی برمی‌گردد. قواعد انجمنی [۱۶] نیازمند تطابق دقیق پیش‌نیازها است و بر علل غالب تمرکز دارد، اما خاموشی‌های گذرا تحت عوامل ناقص و ناپایدار رخ می‌دهند. مدل‌های گروهی LR با مدل‌سازی احتمالاتی و لحاظ ویژگی‌های جزئی، اولویت‌بندی واقع‌گرایانه‌تری ارائه می‌دهند و ENS را مؤثرتر کاهش می‌دهند که ارزش افزوده‌ای برای تعمیرات پیشگیرانه ایجاد می‌کند.

جدول (۶): عملکرد روش پیشنهادی برای شناسایی علل احتمالی خاموشی‌های گذرا، ENS احتمالی قابل پیشگیری و مقایسه عملکرد آنها

با روش پیشنهادی [۱۶]

علت احتمالی خاموشی گذرا	تعداد حضور در اولویت اول	تعداد حضور در اولویت دوم	تعداد حضور در اولویت سوم	مجموع ENS احتمالی	تعداد حضور در اولویت اول	تعداد حضور در اولویت دوم	تعداد حضور در اولویت سوم	مجموع ENS احتمالی
سرکابل	۱۳۸۴	۱۰۰۵	۵۸۰	۳۷۲/۷	۵۲۷۵	۵۶۸	۰	۵۳۹/۱
شرایط بد آب‌وهوایی	۱۲۸۴	۸۴۴	۶۹۰	۲۲۸/۶	۰	۱۱۷	۵۴۹۵	۳۱۱/۲
کابل	۷۴۶	۷۱۶	۵۴۰	۱۶۸/۳	۵۶۸	۵۱۵۸	۱۱۷	۳۱۰/۷
مقره	۲۲۹	۵۸۱	۸۳۰	۶۰/۸	-	-	-	-
خط توزیع هوایی	۲۵۵	۴۴۰	۶۵۲	۵۰/۵	-	-	-	-
پایه	۶۴۶	۶۸۸	۸۶۱	۱۲۰/۳	-	-	-	-
مشترکین	۳۷۰	۶۱۴	۵۹۶	۲۱۲/۹	-	-	-	-
پوشش گیاهی	۷۹۴	۶۵۶	۶۸۳	۵۵/۱	۰	۰	۲۳۱	۵
کات‌اوت فیوز	۴۲	۸	۳	۱۰/۴	-	-	-	-
جامپر	۰	۵	۹	۳/۸	-	-	-	-
ادوات کلیدزنی	۷۰	۱۷۰	۳۱۴	۱۷۵/۳	-	-	-	-
حیوانات	۸۴	۱۵۴	۱۱۲	۱۷/۷	-	-	-	-
ترانسفورماتور	۵	۱۸	۸	۰	-	-	-	-
پست توزیع	۶	۱۰	۱۴	۰/۸	-	-	-	-
برق‌گیر	۰	۶	۲۳	۰	-	-	-	-

روش‌های هوش مصنوعی تفسیرپذیر برای مدیریت داده‌های بزرگ‌تر در پیش گرفته شوند.

۶- نتیجه‌گیری و کارهای آینده

در این مقاله، چارچوبی داده‌محور مبتنی بر مدل‌های گروهی رگرسیون لجستیک برای شناسایی علت خاموشی‌های دائم و گذرا در شبکه توزیع انرژی الکتریکی ارائه شد. ابتدا، با پیش‌پردازش و استخراج ویژگی‌های غنی از داده‌های واقعی (بیش از ۵۰ هزار رکورد خاموشی برنامه‌ریزی‌نشده، سوابق تعمیرات، بار ساعتی فیدرها و اطلاعات آب‌وهوایی)، مدل‌های دودویی برای هر علت آموزش دیده و با رأی‌گیری نرم و وزن‌دهی بر اساس AUC ترکیب گردیدند تا علل محتمل اولویت‌بندی شوند. نتایج دقت بالایی برای اولویت اول مانند ۵۲/۴۶٪ برای حیوانات، ۵۱/۹۲٪ برای کابل و ۷۷/۳۵٪ برای خاموشی‌های گذرا (باوجود تعداد بسیار پایین وقوع برخی از علل خاموشی) نشان داد. همچنین، برای خاموشی‌های گذرا ENS احتمالی ناشی از تبدیل به خاموشی دائم را حدود ۱۹٪ کاهش داد. دستاوردهای اصلی این مقاله شامل افزایش دقت شناسایی هم‌زمان علل دائم و گذرا، تسهیل تعمیرات پیشگیرانه و بهبود شاخص‌های قابلیت اطمینان با استفاده از داده‌های عملیاتی موجود است.

برای پژوهش‌های آینده نیز پیشنهاد می‌شود ادغام داده‌های حسگرهای هوشمند و شکل موج خطا برای غنی‌سازی ویژگی‌ها و بررسی مدل‌های یادگیری عمیق به همراه

سپاسگزاری

این پژوهش در قالب قرارداد پژوهشی و با حمایت شرکت توزیع نیروی برق شیراز انجام شده است. نویسندگان بر خود لازم می‌دانند از همکاری‌های ارزشمند مجموعه شرکت توزیع برق شیراز، به ویژه معاونت بهره‌برداری و دیسپاچینگ، صمیمانه قدردانی کنند.

مراجع

- [1] L. Xu, M.-Y. Chow, "A classification approach for power distribution systems fault cause identification", IEEE Transactions on Power Systems, Vol. 21, No. 1, pp. 53-60, Feb. 2006.
<https://doi.org/10.1109/TPWRS.2005.861981>
- [2] H. Sun, F. Li, C. Sticht, S. Mukherjee, "Outage Cause Classification of Power Distribution Systems with Machine Learning and Real-World Data", in 2022 IEEE Power & Energy Society General Meeting (PESGM), Denver, CO, USA, Jul. 2022.
<http://doi.org/10.1109/PESGM48719.2022.9916833>

- [11] X. Jiang, B. Stephen, S. McArthur, "Automated Distribution Network Fault Cause Identification With Advanced Similarity Metrics", IEEE Transactions on Power Delivery, Vol. 36, No. 2, pp. 785-793, Apr. 2021.
<http://doi.org/10.1109/TPWRD.2020.2993144>
- [12] A. Gholami, A. K. Srivastava, "ORCA: Outage Root Cause Analysis in DER-Rich Power Distribution System Using Data Fusion, Hierarchical Clustering and FP-Growth Rule Mining", IEEE Transactions on Smart Grid, Vol. 15, No. 1, pp. 667-676, Jan. 2024.
<http://doi.org/10.1109/TSG.2023.3281489>
- [13] P. Chawthai, P. Raphisak, S. Katithummarugs, "Association Rule Mining for Power Outages Caused by Animals and Vegetation in Electrical Distribution Systems", in IAIT '23: Proceedings of the 13th International Conference on Advances in Information Technology, Bangkok, Thailand, Dec. 2023.
<https://doi.org/10.1145/3628454.3631857>
- [14] M. Bashkari, A. Sami, M. Rastegar, M. J. Bordbari, "Distribution power system outage diagnosis based on root cause analysis", Scientia Iranica, Vol. 26, pp. 3672-3680, Dec. 2019.
<http://doi.org/10.24200/sci.2019.54190.3638>
- [15] M. R. Dehbozorgi, M. Rastegar, M. Dabbaghjamanesh, "Decision tree-based classifiers for root-cause detection of equipment-related distribution power system outages", IET Generation, Transmission & Distribution, Vol. 14, No. 24, pp. 5809-5815, Dec. 2020. <https://doi.org/10.1049/iet-gtd.2020.0570>
- [16] M. R. Dehbozorgi, M. Rastegar, A. Sami, "Data mining-based cause identification of momentary outages in power distribution systems", Sustainable Cities and Society, Vol. 77, Feb. 2022.
<https://doi.org/10.1016/j.scs.2021.103587>
- [17] Tavanir, "Deputy Directorate for Distribution Coordination – Distribution Supervision Office, Guideline for the Procedure of Recording Outages and Estimating Energy Not Supplied". [In Persian]
- [18] "Renewable.ninja", [Online]. Available: <https://www.renewables.ninja> [Accessed 11 November 2025].
- [19] S. Menard, "Applied Logistic Regression Analysis", SAGE Publications, 1995.
<https://doi.org/10.4135/9781412983433>
- [3] M. R. Dehbozorgi, M. Rastegar, "Association Rule Mining Application to Diagnose Smart Power Distribution System Outage Root Cause", Journal of Energy Engineering and Management, Vol. 11, No. 4, pp. 20-29, Jan. 2022. <http://doi.org/10.22052/11.4.20>
- [4] C. M. Warren, "The effect of reducing momentary outages on distribution reliability indices", IEEE Transactions on Power Delivery, Vol. 7, No. 3, pp. 1610-1617, Jul. 1992.
<http://doi.org/10.1109/TDC.1991.169582>
- [5] M. S. Bashkari, A. Sami, M. Rastegar, "Outage Cause Detection in Power Distribution Systems Based on Data Mining", IEEE Transactions on Industrial Informatics, Vol. 17, No. 1, pp. 640-649, Jan. 2021.
<http://doi.org/10.1109/TII.2020.2966505>
- [6] M. Ebrahimi, M. Rastegar, "A Scalable Ensemble Learning-Based Model for Optimal Placement of Circuit Breaker and Sectionalizer in Power Distribution Systems with the Aim of Reliability Improvement", Computational Intelligence in Electrical Engineering, Vol. 15, No. 3, pp. 1-14, 2024.
<http://doi.org/10.22108/isee.2024.141003.1682> [In Persian]
- [7] M. Haghshenas, R. Hooshmand, "Resilience-Oriented Power Distribution System Planning and Outage Management Based on Operational Situation Awareness", Computational Intelligence in Electrical Engineering, Vol. 16, No. 1, pp. 15-30, 2025.
<http://doi.org/10.22108/isee.2025.144861.1729>
- [8] M. Doostan, R. Sohrabi, B. Chowdhury, "A data-driven approach for predicting vegetation-related outages in power distribution systems", International Transactions on Electrical Energy Systems, 2020. <https://doi.org/10.1002/2050-7038.12154>
- [9] D. Cerrai, P. Watson, E. N. Anagnostou, "Assessing the effects of a vegetation management standard on distribution grid outage rates", Electric Power Systems Research, Vol. 175, Oct. 2019.
<http://doi.org/10.1016/j.epsr.2019.105909>
- [10] M. Doostan, B. Chowdhury, "Predicting Lightning-Related Outages in Power Distribution Systems: A Statistical Approach", IEEE Access, Vol. 8, pp. 84541-84550, 2020.
<http://doi.org/10.1109/ACCESS.2020.2991923>

- [22] W. J. Youden, "Index for rating diagnostic tests", *Cancer*, Vol. 3, No. 1, p. 32–35, 1950.
[https://doi.org/10.1002/1097-0142\(1950\)3:1<32::aid-cnrcr2820030106>3.0.co;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cnrcr2820030106>3.0.co;2-3)
- [20] Z.-H. Zhou, "Ensemble Methods: Foundations and Algorithms", Taylor & Francis, 2012.
<https://doi.org/10.1201/b12207>
- [21] H. He, E. A. Garcia, "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 9, pp. 1263-1284, 2009.
<https://doi.org/10.1109/tkde.2008.239>

¹ System Average Interruption Duration Index (SAIDI)

² System Average Interruption Frequency Index (SAIFI)

³ Association Rules

⁴ Artificial Neural Network (ANN)

⁵ Logistic Regression-LR

⁶ Ensemble Learning with Resampling

⁷ Soft Voting

⁸ Area Under the Curve (AUC)

⁹ Energy Not Supplied (ENS)

¹⁰ Binary Classification

¹¹ Ensemble Models

¹² One-vs-Rest Classification

¹³ Meta-Learner

¹⁴ Soft Voting

¹⁵ Youden's J statistic

¹⁶ Anomaly Detection

¹⁷ Isolation Forest

¹⁸ Hyperparameter

¹⁹ Grid Search

²⁰ Hard-Voting

²¹ Inference

²² Robustness

²³ Cost Matrix

