بهبود دقت تشخیص سرقت برق در شبکه هوشمند

رضا الیاسی[۱]، کمال جمشیدی[۲]، علی بهلولی[۳]

[۱]دانشگاه اصفهان، r.elyasi@eng.ui.ac.ir

[۲]دانشگاه اصفهان، jamshidi@eng.ui.ac.ir

[۳]دانشگاه اصفهان، bohlooli@eng.ac.ir

چکیده

شبکه‌های هوشمند برق به دلیل به‌کارگیری زیرساخت‌های کامپیوتری و مخابرات دیجیتال، در مقایسه با شبکه‌های سنتی در برابر حملات امنیتی آسیب‌پذیرترند. به همین دلیل نگرانی‌هایی در مورد نفوذ افراد غیرمجاز به این تجهیزات به‌منظور سرقت برق و اعمال خرابکارانه در شبکه برق وجود دارد. در این میان سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری می‌توانند به عنوان روشی برای تشخیص افراد فریب‌کار به کار روند. روش‌های پیشین ارائه شده برای تشخیص سرقت برق دارای مشکلاتی از قبیل دقت تشخیص پایین، هزینه پیاده‌سازی و نرخ مثبت کاذب زیاد هستند. هدف این مقاله، بهبود عملکرد سیستم تشخیص نفوذ در تشخیص حملات مربوط به تغییر داده در شبکه هوشمند است. به منظور دستیابی به هدف مذکور، ابتدا چالش‌های امنیتی پیش روی شبکه هوشمند شناسایی و معرفی می‌گردند و در ادامه روشی مبتنی بر یادگیری ماشین ارائه می‌شود. روش پیشنهادی از دو بخش تشکیل شده است. در بخش نخست موتور تشخیص حملات به کمک چهار دسته‌بند پایه‌ی ماشین بردار پشتیبان، شبکه عصبی MLP، درخت تصمیم و K- نزدیک‌ترین همسایه پیاده‌سازی می‌شود و در بخش دوم نتایج این دسته‌بندها به کمک الگوریتم تلفیقی AdaBoost بهبود می‌یابد. روش پیشنهادی با استفاده از نرم‌افزار MATLAB 2013a پیاده‌سازی و با مجموعه داده CER ارزیابی شد. نتایج حاصل از ارزیابی نشان می‌دهد که دقت تشخیص و نرخ مثبت کاذب این روش در مقایسه با سایر دسته‌بندها به ترتیب به ٪۸۶/۷۱ و ٪۱۳/۳ بهبود یافته است.

واژه‌های کلیدی

شبکه هوشمند، سرقت برق، سیستم تشخیص نفوذ، زیرساخت اندازه گیری پیشرفته، الگوریتم AdaBoost

۱- مقدمه

امروزه صنعت برق را می‌توان یکی از پایه‌های اصلی اقتصاد و صنعت هر کشوری دانست. با پیشرفت و رشد فناوری، صنایع و جمعیت، میزان نیاز به انرژی الکتریکی به‌طور قابل ملاحظه‌ای افزایش یافته است. در حال حاضر شبکه برق کارایی بالایی ندارد. از شاخص‌های مهم ناکارآمدی شبکه‌های سنتی برق می‌توان موارد زیر را نام برد[۱]:

- عدم کار آیی شبکه برق در مدیریت حداکثر تقاضا.
- عدم توانایی شبکه در ایجاد تبادل مطمئن اطلاعات.
- قابلیت محدود شبکه در استفاده از منابع تولید پراکنده.
- ناکارآمدی شبکه با گسترش اتصال خودروهای الکتریکی.
- مستعد بودن شبکه در بروز خاموشی و اختلال کیفیت توان.

وجود چنین مشکلاتی در شبکه برق سنتی موجب شکل گرفتن چهره جدیدی از شبکه برق به نام شبکه هوشمند (SG)[۱] گردید. شبکه هوشمند، یک شبکه قدرت پیشرفته است که از فناوری اطلاعات و ارتباطات (ICT)[۲] برای اداره کردن، پایش کردن، کنترل کردن داده‌های مربوط به فعالیت‌های تهیه‌کنندگان و مشتریان برای رسیدن به قابلیت اطمینان بالا، انعطاف‌پذیری، کارایی و سود بیشتر استفاده می‌کند. شبکه هوشمند، حاصل به‌کارگیری فناوری‌های مخابراتی در شبکه الکتریکی است که امکان ارتباط دو طرفه بین شرکت‌های برق را فراهم می‌کند.

با وجود مزایای فراوان و تسهیلات منحصر به فرد شبکه هوشمند، این شبکه با چالش‌های جدیدی مانند در دسترس بودن، تحمل خطا، امنیت روبرو است. سیستم‌های هوشمند برق در مقایسه با شبکه سنتی در برابر حملات امنیتی آسیب‌پذیرترند، چرا که آن‌ها به شبکه‌های بی‌سیم، اینترنت، ارتباطات موبایل و فرکانس رادیویی (RF) متصل‌اند. این شبکه مخابراتی باعث ایجاد ارتباط

[1] Smart Grid

[2] Information and Communication Technology

بین اجزای مختلف سیستم خواهد شد و امکان برقراری ارتباط دوطرفه را فراهم خواهد نمود. بستر مخابراتی نقاط دسترسی زیادی برای نفوذ به اجزای مختلف شبکه ایجاد می‌نماید و همین موضوع باعث به خطر افتادن امنیت شبکه خواهد شد. در شبکه هوشمند میلیون‌ها کنتور هوشمند وجود دارد که به‌طور غیرمستقیم حامل اطلاعات خصوصی و محرمانه از مشتریان خود خواهند بود. اطلاعات مصرف انرژی مشتریان در کنتور ذخیره شده و از طریق کانال مخابراتی به سمت مرکز داده‌ها ارسال می‌گردد. این اطلاعات مصرف می‌تواند نشان دهنده رفتار و عادت روزانه، هفتگی و یا ماهیانه مشتریان باشد. مشتریان فریبکار می‌توانند با استفاده از آسیب‌پذیری‌های شبکه مخابراتی و کنتورهای هوشمند، میزان انرژی مصرفی قرائت شده توسط کنتورها را تغییر داده و به تبع آن باعث کاهش هزینه مصرفی شوند. همچنین سارقان می‌توانند با دست‌یابی به الگوی مصرف یک مشتری به زمان خالی بودن، زمان خواب و سایر رفتارهای مشتری پی ببرد و از آن جهت انجام سرقت استفاده کنند. از این رو کنتورهای هوشمند موضوع جذابی برای افراد بداندیش می‌باشند. تحقیقات نشان می‌دهد که شرکت‌های توزیع برق در جهان سالانه بیش از ۲۵ میلیارد دلار بابت سرقت برق زیان می‌بینند. شرکت‌های برق هند به عنوان یک نمونه از قربانیان سرقت برق، سالانه در حدود ۴/۵ میلیارد دلار از این طریق از دست می‌دهند که تنها با بازیابی ۱۰٪ از این مقدار می‌توانند GWh ۸۳۰۰۰ انرژی ذخیره کنند و از انتشار ۹/۲ میلیون تن کربن دی اکسید بکاهند[۲]. همچنین طبق گزارش منتشر شده توسط اداره تحقیقات آمریکا (FBI)[1] در سال ۲۰۱۰، حمله‌ای با هدف سرقت برق علیه AMI انجام گرفت که طی آن شرکت‌های برق متحمل زیان ۴۰۰ میلیون دلاری شدند[۳].

روش‌های تشخیص سرقت برق به عمدتا سه دسته‌ی مبتنی بردسته‌بندی، مبتنی بر حالت و مبتنی بر نظریه بازی‌ها طبقه‌بندی می‌شوند [۴]:

روش‌های مبتنی بر دسته‌بندی از داده‌های مصرف انرژی جمع‌آوری شده از طریق AMI بهره می‌گیرند. میزان مصرف مشترک در شرایط عادی از یک الگوی آماری خاصی پیروی می‌کند اما هرگونه ناهنجاری در مصرف می‌تواند نشانه‌ای از فعالیت‌های بدخواهانه باشد که این ناهنجاری منجر به تغییر در الگوی مصرف مشترک می‌شود. روش‌های داده کاوی و یادگیری ماشین از جمله روش‌هایی هستند که با استفاده از مجموعه داده‌ها به دسته‌بندی مشتریان درست‌کار و فریب‌کار می‌پردازند. روش‌های تشخیص مبتنی بر حالت از تجهیزات خاصی مانند حسگرهای بی‌سیم و شناسه‌های RFID استفاده می‌کنند که نرخ مثبت کاذب را کاهش و دقت تشخیص را بهبود می‌بخشند. این روش‌ها برای پایش کردن سیستم نیازمند هزینه‌های اضافی از قبیل هزینه تجهیزات، هزینه پیاده‌سازی سیستم، هزینه نرم‌افزاری، هزینه اجرا و آموزش هستند. در روش مبتنی بر نظریه بازی‌ها، مسئله تشخیص سرقت برق به صورت یک بازی بین سارق و شرکت برق فرمول‌بندی می‌شود. با وجود این که این روش کم هزینه و معقول است اما راه حلی بهینه برای تشخیص سرقت برق نیست. همچنین چگونه فرمول‌بندی کردن تابع مطلوبیت[2] تمام بازیکن‌ها از قبیل سارقین و بخش توزیع نیز یک مسئله چالش برانگیز است.

جوکار و همکاران [۵] روشی دو مرحله‌ای، مبتنی بر الگوی مصرف برای تشخیص سرقت برق در زیرساخت اندازه‌گیری پیشرفته (AMI) شبکه هوشمند ارائه دادند. مرحله اول با استفاده از الگوریتم خوشه‌بندی k-mean داده‌های مربوط به میزان برق مصرفی هر مشترک خوشه‌بندی می‌گردد. سپس این داده‌های خوشه‌بندی شده با استفاده از الگوریتم دسته‌بندی SVM[3] تک کلاس و چند کلاس، به دو دسته‌ی نرمال و غیر نرمال دسته‌بندی می‌شوند. در مرحله دوم با استفاده از نتایج حاصل از دسته‌بندی SVM و میزان برق مصرفی در ترانسفورماتور و کنتور هوشمند الگوریتمی پیشنهاد گردید که با حفظ حریم خصوصی مشتریان، دقت تشخیص را بهبود می‌دهد. آنجلوس و همکاران [۶] یک روش محاسباتی برای دسته‌بندی پروفایل مصرف برق پیشنهاد دادند. روش پیشنهادی شامل دو بخش می‌باشد. در بخش اول با استفاده از الگوریتم خوشه‌بندی C-mean فازی، مشتریان با پروفایل مشابه در یک خوشه قرار می‌گیرند. در بخش دوم دسته‌بندی فازی انجام شده و فاصله اقلیدسی تا مرکز خوشه اندازه‌گیری می‌شود. سپس فاصله‌های اندازه‌گیری شده نرمال‌سازی و مرتب می‌شوند و مشتریانی که فاصله زیادی با مرکز خوشه دارند به عنوان مشتریان متقلب دسته‌بندی می‌شوند. نگی و همکاران [۷] روشی برای تشخیص تلفات غیر فنی مبتنی بر ماشین بردار پشتیبان ارائه دادند. سپس ماشین بردار پشتیبان با استفاده از اطلاعات پروفایل برق مصرفی هر مشترک و برخی ویژگی‌های دیگر ، هر گونه ناهنجاری در میانگین مصرف را تشخیص می‌دهد هرگونه ناهنجاری در میانگین مصرف را تشخیص می‌دهد و مشتریان را به دو دسته درست‌کار و فریب‌کار دسته‌بندی می‌کند. پارامترهای SVM با استفاده از جستجوی گرید بهینه‌سازی می‌شوند. نتایج حاصل از روش پیشنهادی نشان می‌دهد که این روش نرخ تشخیص را از ۳٪ فعلی به ۶۰٪ افزایش می‌دهد. نویسندگان در ادامه تحقیقات خود از الگوریتم ژنتیک و منطق فازی برای بهبود روش اولیه استفاده نمودند [۸ و ۹]. در [۱۰] روشی مبتنی بر درخت تصمیم و SVM پیشنهاد شد و در [۱۱] نیز روشی مبتنی بر نظریه بازی‌ها ارائه گردید.

با توجه به اهمیت موضوع سرقت برق در شبکه‌های توزیع هوشمند که منجر به ایجاد مشکلات فنی و ضررهای اقتصادی می‌شود نیازمند روشی مطمئن برای شناسایی مشترکین متقلب هستیم. این مقاله در چهار بخش تنظیم شده است. در بخش دوم روش پیشنهادی شرح داده می‌شود. در بخش سوم روش ارائه شده پیاده سازی و ارزیابی می‌شود و نهایتاً در بخش چهارم نتایج به دست آمده بیان شده و مورد بحث قرار می‌گیرد و پیشنهادهایی برای ادامه کار در این زمینه ارائه می‌گردد.

---

[1] **Federal Bureau Investigation**

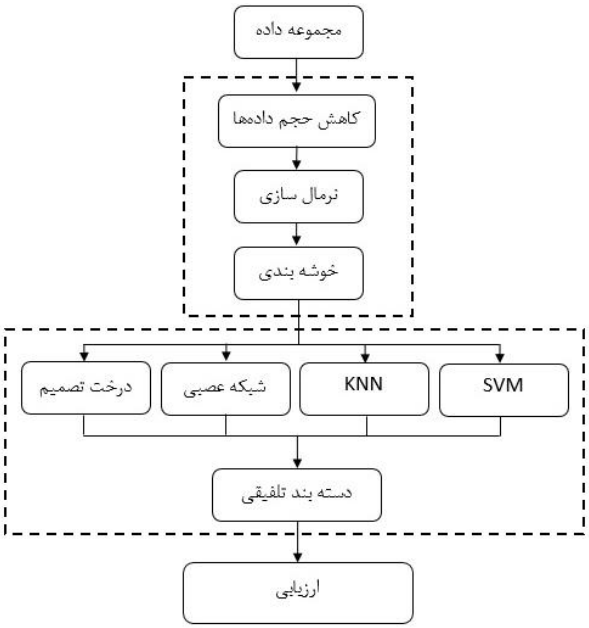
[2] **Utility Function**

[3] **Support Vector Machine**

۲- روش پیشنهادی

روش پیشنهادی در این پژوهش متشکل از سه مرحله می‌باشد. این مراحل عبارت‌اند: ۱. جمع آوری داده. ۲. عملیات پیش پردازش ۳. استفاده از دسته‌بند تلفیقی.

چارچوب کلی طرح پیشنهادی مطابق شکل ۱ است. هر یک از مراحل فوق در ادامه با جزئیات توضیح داده خواهند شد.



شکل ۱: نمای کلی روش پیشنهادی

۲-۱ جمع آوری داده

معمولاً از داده‌های تاریخی که سابقه مصرف برق هر مشتری را نشان می‌دهند برای تشخیص سرقت برق استفاده می‌شود. کنتورهای هوشمند، میزان مصرف برق را چندین مرتبه در طول روز اندازه‌گیری و به شرکت برق ارسال می‌کنند. بنابراین برای هر مشترک در شرکت برق یک پروفایل مصرف وجود دارد که میزان مصرف برق روزانه در طی ساعات مختلف در آن ثبت شده است. این داده‌ها به صورت یک بردار به پیمانه کاهش حجم داده ها ارسال می‌شود. با استفاده از این پروفایل می‌توان الگوی مصرف برق هر مشترک را استخراج کرد و با استفاده از روش پیشنهادی بر روی آن تحلیل کرد.

۲-۲ عملیات پیش پردازش

گام بعدی پیش‌پردازش داده‌ها است که منجر به افزایش کیفیت داده‌های مورد استفاده می‌شود. این مرحله شامل سه نوع عملیات است که عبارت‌اند از کاهش حجم داده‌ها، نرمال‌سازی داده‌ها و خوشه‌بندی داده‌ها.

۲-۲-۱ نرمال سازی داده ها

فرآیند نرمال‌سازی عبارت است از تغییر مقیاس داده‌ها به گونه‌ای که داده‌ها به یک فاصله‌ی کوچک و معین نگاشت شوند. این عمل باعث می‌شود که داده‌ها با مقیاس بزرگ نتایج را به سمت خود منحرف نکنند و دقت افزایش یابد. در این پژوهش از روش min-max استفاده می‌گردد. نرمال‌سازی min-max یک تبدیل خطی مطابق با فرمول ۱ را بر روی داده‌های اولیه اجرا می‌کند و داده‌ها به بازه‌ی بین صفر تا یک نگاشت می‌شوند.

$$\acute{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (۱)$$

۲-۲-۲ کاهش حجم داده ها

هر بردار داده در مجموعه داده شامل میزان مصرف برق مشتری طی دوره ۲۴ ساعت است، برای مثال اگر در هر ساعت n بار عمل اندازه‌گیری مصرف برق انجام شود آنگاه بردار داده شامل $24 \times n$ نمونه می‌شود. روش‌های مختلفی برای کاهش حجم داده‌ها وجود دارد که با استخراج ویژگی‌های مهم داده، اتلاف اطلاعات را حداقل می‌کند. از روش نمونه‌برداری می‌توان به‌عنوان یک روش برای کاهش داده‌ها استفاده کرد، زیرا با کمک آن می‌توان مجموعه‌ی بزرگی از داده‌ها را با کمک نمونه‌های تصادفی و بسیار کمتری نشان داد. نرخ نمونه‌گیری ارتباط مستقیمی با حریم خصوصی مشتری دارد. هرچه نرخ نمونه‌برداری بیشتر باشد آنگاه ریسک افشای اطلاعات شخصی افراد بیشتر می‌شود. بنابراین باید نرخ نمونه‌برداری را به‌گونه‌ای انتخاب کرد تا علاوه بر دقت تشخیص بالای مشتریان درست‌کار و فریب‌کار، حریم خصوصی مشتریان را نیز حفظ کند. در این پژوهش برای کاهش حجم، داده ها با نرخ‌های متفاوت به‌صورت چند در میان نمونه‌گیری می‌شوند یعنی در مجموعه داده، میزان مصرف برق در هر نیم ساعت اندازه‌گیری شده است بنابراین هر بردار داده دارای ۴۸ نمونه است. این نمونه‌ها به صورت چند در میان انتخاب می‌شوند و حجم مجموعه داده‌ها را کاهش می‌دهند.

۲-۲-۳ خوشه بندی

خوشه بندی به فرآیندی اطلاق می‌شود که در آن مجموعه‌ای از اشیا به چندین دسته یا خوشه گروه‌بندی می‌شوند، به ترتیبی که داده‌های درون یک خوشه بسیار شبیه به یکدیگر و اشیا خوشه‌های مختلف بسیار متفاوت هستند. در این پژوهش از الگوریتم خوشه‌بندی k-means که یک نوع خوشه‌بندی افرازی است، استفاده می‌شود. برای به دست آوردن نرخ تشخیص بهتر، الگوریتم خوشه‌بندی k-mean با مقادیر مختلف k مجموعه داده را خوشه‌بندی می‌کند. این الگوریتم به ترتیب زیر عمل می‌کند:

- با استفاده از نمودار سیلوئت تعداد خوشه بهینه محاسبه و k شی را به منزله مراکز خوشه‌های ابتدایی انتخاب می‌کند.
- هر شی را با توجه به بیشترین شباهت آن به مراکز خوشه‌ها، به خوشه‌ها تخصیص می‌دهد.
- مراکز خوشه‌ها را به روز می‌کند به این معنا که برای هر خوشه مقدار متوسط اشیای آن خوشه را محاسبه می‌کند.
- تا هنگامی که هیچ تغییری در خوشه‌ها رخ ندهد، به مرحله دوم رجوع می‌کند.

خوشه‌هایی که تعداد اعضای کمی دارند حذف می‌شوند و برای آموزش دسته‌بند استفاده نخواهند شد. این کار باعث جلوگیری از آلوده شدن مجموعه داده نرمال توسط حملات ناشناخته خواهد شد.

۲-۳ دسته‌بند تلفیقی

در روش تلفیقی از k دسته‌بند پایه $M_1, M_2, \dots, M_K$ برای ایجاد مدل دسته‌بندی ترکیبی M^* که بهبود یافته است، استفاده می‌شود. از مجموعه داده‌های D برای تولید k مجموعه‌ی آموزشی $D_1, D_2, \dots, D_K$ استفاده می‌شود و با کمک هر یک از آن‌ها دسته‌بندهای $M_i\ (1 \le i \le k)$ نیز

خلق می‌شوند. برای دسته‌بندی یک نمونه جدید هر یک از دسته‌بندهای پایه پیش‌گویی کلاس خود را به عنوان یک رأی بر می‌گردانند و روش بر اساس آرا جمع‌آوری شده، رأی نهایی یعنی برچسب کلاس را تعیین می‌کند[۱۲]. از میان روش‌های boosting الگوریتم AdaBoost یکی از الگوریتم‌های رایج محسوب می‌شود. مجموعه داده D از تعداد d نمونه که هر یک به شکل (X_i, y_i) هستند تشکیل شده است و در آن برچسب کلاس نمونه X_i با y_i نشان داده می‌شود. در ابتدا الگوریتم AdaBoost مقدار وزن یکسان $1/d$ را برای هر نمونه آموزشی در نظر می‌گیرد. در این روش تلفیقی برای تولید k دسته‌بند به k مرحله نیاز است. در مرحله iام با کمک نمونه‌گیری با جایگزینی مجموعه داده D_i شکل می‌گیرد. همان‌طور که می‌دانید در این روش ممکن است نمونه‌های یکسانی بیش از یک‌بار انتخاب شوند. شانس انتخاب نمونه بر اساس وزن نمونه تعیین می‌شود. با استفاده از مجموعه داده‌های آموزشی D_i مدل M_i ساخته می‌شود و با کمک همین مجموعه، خطای مدل نیز محاسبه می‌گردد. پس از آن وزن‌های نمونه آموزشی بر اساس این که چگونه دسته‌بندی شده‌اند، تنظیم می‌شود

در این روش اگر نمونه‌ای به درستی دسته‌بندی نشده باشد، وزن آن افزایش می‌یابد و وزن نمونه‌هایی که به درستی دسته‌بندی شده‌اند کاهش داده می‌شود. در واقع وزن یک نمونه، دشواری دسته‌بندی آن نمونه را منعکس می‌کند، به صورتی که وزن بالاتر نشان می‌دهد اغلب این نمونه به درستی دسته‌بندی نشده است. وزن‌های جدید نمونه‌ها برای تولید مجموعه آموزشی دسته‌بندی بعدی در مرحله‌ی بعد استفاده می‌شود. ایده‌ی اصلی این کار از جایی سرچشمه می‌گیرد که ما هنگام ساخت دسته‌بند مایلیم که دسته‌بند پایه بیشتر بر روی دسته‌بندی نمونه‌هایی تمرکز کند که در مرحله قبلی به درستی دسته‌بندی نشده‌اند. ممکن است برخی از دسته‌بندها در دسته‌بندی نمونه‌های دشوار بهتر از بعضی دیگر عمل کنند. بدین طریق یک سری از دسته‌بندهایی ساخته شده‌اند که یکدیگر را تکمیل می‌کنند. شبه کد الگوریتم AdaBoost به صورت زیر است [۱۲].

AdaBoost Algorithm:

Input:

- ***D*** , a set of d − class − labeled training tuples;
- ***k***, the number of round (one classifier is generated per round);
- a classification learning schem;

Output:

1. initialize the weight of each tuple in D to $1/d$;
2. **for** $i = 1$ to k **do** //for each round;
3. sample D With replacement according to the tuple weight to obtain D_i;
4. use training set D_i to derive a model, M_i;
5. compute $error(M_i)$, the error rate of $M_i = \sum_{j=1}^{d} w_j \times err(X_j);|$
6. **if** $error(M_i) > 0.5$ **then**
7. go back to step 3 and try again;
8. **endif**
9. **for** each tuple in D_i that was correctly classified **do**
10. multiply the weight of the tuple by $error(M_i)/(1 - error(M_i))$; //update weight
11. normalize the weight of each tuple;
12. **endfor**

شکل ۲: شبه کد الگوریتم AdaBoost

در این پژوهش از چهار دسته‌بند شبکه عصبی MLP، ماشین بردار پشتیبان، درخت تصمیم و K- نزدیک‌ترین همسایه به عنوان دسته‌بندهای پایه استفاده شده است.

۳- شبیه سازی و ارزیابی

در این بخش به کمک شبیه‌سازی روش پیشنهادی ارزیابی می‌شود. کلیه شبیه‌سازی‌ها به وسیله رایانه با واحد پردازش مرکزی Core i5 و حافظه 4GB و در محیط نرم‌افزار MATLAB R2013a انجام شده است.

۴-۱ مجموعه داده

مجموعه داده‌ای که در این پژوهش استفاده شده، به عنوان بخشی از یک آزمایش برای مطالعه فناوری‌های ارتباطی کنتورهای هوشمند توسط کمیسیون تنظیم انرژی ایرلند (CER) جمع‌آوری گردیده است. در این مجموعه داده میزان مصرف برق ۶۰۴۸ مشتری در بازه زمانی بین سال‌های ۲۰۰۹ تا ۲۰۱۰ (۷۴ هفته) ثبت شده است. در این مجموعه داده میزان مصرف برق هر مشتری در هر نیم ساعت توسط کنتور هوشمند ثبت شده است. هر نمونه از این مجموعه داده دارای سه ویژگی به صورت زیر است:

جدول ۱: ویژگی‌های هر سطر از مجموعه داده

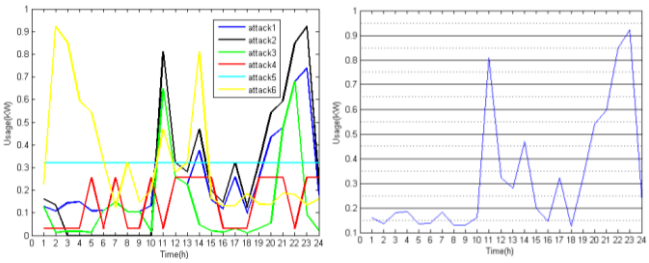
میزان مصرف	کد	شناسه
۰/۱۴	۱۹۵۰۱	۱۳۹۲

مشتریانی که در این آزمایش شرکت کردند، کنتورهای هوشمند در منازلشان نصب گردید بنابراین از لحاظ منطقی، نمونه‌های اندازه‌گیری شده متعلق به افراد درست‌کار است و احتمال این که افراد شرکت‌کننده در آزمایش مرتکب سرقت شوند بسیار کم است، بنابراین داده‌های حمله در این مجموعه داده وجود ندارد. روشی که در اینجا استفاده شده است، ایجاد یک مجموعه داده غیر نرمال با استفاده از مجموعه داده نرمال است. هدف از سرقت برق آن است که مشتری مصرف برق کمتری نسبت به میزان واقعی گزارش دهد و یا مصرف زیاد را به دوره‌های با تعرفه پایین شیفت دهد. با مطالعه سناریوهای مختلف شش نوع تابع به صورت جدول ۲ تعریف شده است که حملات ممکن را مدل‌سازی می‌کند:

جدول ۲: توابع تولید کننده حمله در الگوی مصرف برق [۵].

۱	$h_1(x_t) = \alpha x_t \ , \quad \alpha = random(0.1, 0.8);$
۲	$h_2(x_t) = \beta_t x_t$ $\beta_t = \begin{cases} 0 & start\ time \le t \le end\ time \\ 1 & else \end{cases}$ start time = random(0, 23 − minOffTime) duration = random(minOffTime, 24) end time = start time + duration here minOffTime = 4;
۳	$h_3(x_t) = \gamma_t x_t \ , \quad \gamma_t = random(0.1, 0.8);$
۴	$h_4(x_t) = \gamma_t\, mean(x), \ \gamma_t = random\ (0.1,0.8)$
۵	$h_5(x_t) = mean(x);$
۶	$h_6(x_t) = x_{24-t}$

شکل ۳ (الف) نمودار مصرف برق یک مشتری را در طول یک روز نشان می‌دهد و شکل ۳ (ب) نمودار متناظر آن را با اعمال حملات فوق نشان می‌دهد. الگوی مصرف برق مشتریان در حالت معمول و حالت حمله تفاوت دارد، بنابراین به کمک این الگوها می‌توان مشتریان متقلب را شناسایی کرد.



شکل ۳: (الف) نمودار الگوی مصرف برق روزانه. (ب) نمودار الگوی مصرف برق روزانه بعد از اعمال حملات.

۴-۲ ارزیابی

در این پژوهش برای نشان دادن کارایی روش پیشنهادی از پنج معیار ارزیابی مختلف استفاده می‌شود. این معیارها شامل صحت[1] (نرخ تشخیص)، دقت[2]، نرخ مثبت کاذب، باز فراخوانی[3] و F-measure می‌شوند. برای ارزیابی روش پیشنهادی، ابتدا ۱۰۰۰ مشترک خانگی از مجموعه داده انتخاب گردید. با توجه به این که نرخ نمونه برداری تاثیر مستقیمی بر روی کارایی دارد، روش پیشنهادی با نرخ های نمونه برداری مختلف ارزیابی شد.

- نرخ تشخیص: درصد نمونه‌هایی از مجموعه داده که برچسب آن‌ها به درستی توسط سیستم تعیین می‌شود. به این معیار ارزیابی نرخ صحت نیز گفته می‌شود. نحوه محاسبه صحت در رابطه ۲ آمده است.

$$Accuracy = \frac{TP + TN}{P + N} \qquad (۲)$$

جدول ۳: نتایج ارزیابی معیار نرخ تشخیص

	Accuracy				
الگوریتم	N=24	N=12	N=10	N=6	N=2
SVM	74.25	80.87	82.09	76.54	66.22
MLP	76.17	83.18	83.27	77.54	66.18
K-NN	70.11	74.72	74.97	71.40	64.05
Tree	74.94	80.30	81.36	75.99	65.09
AdaBoost	74.47	85.47	86.71	80.57	67.63

- باز فراخوانی: این معیار نشان می‌دهد که چه درصدی از نمونه‌های نرمال، توسط سیستم نیز نرمال در نظر گرفته شده‌اند. نحوه محاسبه Recall در رابطه ۳ آمده است.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{P} \qquad (۳)$$

جدول ۴: نتایج ارزیابی معیار بازفراخوانی

	Recall				
الگوریتم	N=24	N=12	N=10	N=6	N=2
SVM	77.28	79.98	81.98	79.67	82.98
MLP	83.83	90.14	90.20	85.95	87.29
K-NN	73.43	75.76	76.43	74.54	73.05
Tree	75.79	83.44	83.53	78.25	68.95
AdaBoost	83.33	89.26	91.70	86.80	82.58

- F-measure: این معیار ارزیابی به گونه‌ای دو معیار باز فراخوانی و دقت را با یکدیگر ترکیب می‌کند و آن‌ها را در قالب یک معیار نشان می‌دهد. در واقع F-measure نشان دهنده یک میانگین هارمونیک از دو معیار دقت و باز فراخوانی نیز می‌باشد. نحوه محاسبه F-measure در رابطه ۴ آمده است.

$$F - measur = \frac{2 * precision * recall}{precision + recall} \qquad (۴)$$

جدول ۵: نتایج ارزیابی معیار F-measure

	F-measure				
الگوریتم	N=24	N=12	N=10	N=6	N=2
SVM	78.57	82.23	83.83	80.18	80.54
MLP	82.11	86.39	86.55	82.82	81.85
K-NN	74.93	78.91	78.61	75.59	74.49
Tree	79.65	85.24	85.91	80.88	72.66
AdaBoost	83.58	87.92	89.40	85.11	80.91

- نرخ مثبت کاذب: این معیار نشان می‌دهد که چه درصدی از نمونه‌های نرمال، به اشتباه توسط سیستم غیر نرمال تشخیص داده می‌شود. نحوه محاسبه نرخ مثبت کاذب در رابطه ۵ آمده است.

$$FPR = \frac{FP}{FP + TN} \qquad (۵)$$

جدول ۶: معیار ارزیابی نرخ مثبت کاذب

	FPR				
الگوریتم	N=24	N=12	N=10	N=6	N=2
SVM	18.55	14.03	13.27	18.57	20.34
MLP	18.90	18.45	17.84	20.99	25.44
K-NN	23.53	19.99	19.98	24.01	27.00
Tree	13.91	11.93	10.59	14.84	18.03
AdaBoost	15.67	13.70	13.37	16.94	19.63

- دقت: این معیار نشان می‌دهد که چه درصدی از نمونه‌های نرمال، توسط سیستم نیز مثبت در نظر گرفته شده‌اند. نحوه محاسبه دقت در رابطه ۶ آمده است.

$$precision = \frac{TP}{TP + FP} \qquad (۶)$$

[1] **Accuracy**

[2] **Precision**

[3] **Recall**

جدول ۷: نتایج ارزیابی معیار دقت

	Precision				
الگوریتم	N=24	N=12	N=10	N=6	N=2
SVM	80.87	85.28	86.23	81.33	80.62
MLP	81.89	83.15	83.51	80.48	78.02
K-NN	76.67	79.54	79.63	76.08	73.24
Tree	84.59	87.63	88.76	84.58	79.85
AdaBoost	84.36	86.84	87.33	83.8	81.38

به منظور مطالعه تأثیر نرخ نمونه برداری بر روی کارایی، معیارهای ارزیابی با نرخ نمونه برداری متفاوت آزمایش شده است. مشاهده گردید که بیشترین کارایی با نرخ ۱۰ نمونه در روز به دست می‌آید که نشان می‌دهد نرخ نمونه برداری بالاتر لزوماً کارایی بهتری را ایجاد نمی‌کند. این مسئله ممکن است به دلیل این واقعیت باشد که برای نرخ‌های نمونه برداری پایین‌تر، تأثیر سرقت بر روی هر نمونه ممکن است واضح‌تر باشد. برای مثال زمانی که یک مشتری متقلب بخواهد 5KW کمتر از میزان مصرف واقعی‌اش به شرکت برق گزارش دهد، با تقسیم یکنواخت این مقدار در ۲۴ نمونه، تغییر الگوی مصرف کم است اما اگر این میزان در ۱۰ نمونه تقسیم شود آنگاه تغییر الگوی مصرف چشم‌گیر است و دقت تشخیص بالاتر می‌رود. بنابراین ممکن است نرخ نمونه برداری پایین‌تر، دقت تشخیص را افزایش دهد. از طرف دیگر نرخ نمونه برداری بسیار پایین باعث می‌شود تا نتوان به صورت دقیق الگوی مصرف را مدل‌سازی کرد. بنابراین انتخاب کمترین نرخ نمونه برداری که کارایی قابل قبولی را فراهم کند بسیار مهم است زیرا نه تنها منابع مورد نیاز را حداقل می‌کند بلکه حریم خصوصی مشتریان را حفظ می‌کند. در [۱۳] نشان داده شد که با نرخ نمونه برداری ثانیه/نمونه ۰,۵ یک شخص می‌تواند اطلاعات جزئی از قبیل عنوان فیلم یا کانال تلویزیونی در حال پخش در منزل را متوجه شود. در [۱۴] نشان داده شد که با نرخ نمونه برداری دقیقه/نمونه ۰,۵ یک شخص می‌تواند نوع و زمان استفاده از لوازم خانگی را متوجه شود. با نرخ نمونه برداری روز/نمونه ۱ یک شخص می‌تواند اطلاعاتی با جزئیاتی در حد این که آیا مشترک در منزل حضور دارد یا نه به دست آورد. به عنوان مثال از این اطلاعات می‌توان برای انجام سرقت منزل استفاده کرد. هرچه نرخ نمونه‌برداری کمتر باشد اطلاعات کمتری از داده‌های اندازه‌گیری شده می‌تواند استخراج گردد. نتایج بدست آمده نشان دهنده‌ی بهبود دقت و عملکرد روش پیشنهادی نسبت به سایر روش‌های دسته‌بندی است. همچنین روش پیشنهادی از نظر میزان نیاز به نمونه‌برداری نسبت به روش‌های دیگر دارای مقدار کم‌تری است که باعث حفظ حریم خصوصی مشتریان می‌شود.

۴- نتیجه‌گیری و پیشنهادها

شبکه هوشمند نسل جدیدی از شبکه‌های توزیع برق هستند که از بکارگیری فناوری ارتباطات و اطلاعات در سیستم قدرت حاصل می‌شود. شبکه هوشمند به دلیل بکارگیری زیرساخت‌های کامپیوتری و مخابرات دیجیتال، در مقایسه با شبکه سنتی آسیب‌پذیرتر است و نگرانی‌هایی در مورد نفوذ افراد غیرمجاز به این تجهیزات به منظور سرقت برق وجود دارد. در این مقاله روشی برای تشخیص سرقت برق در شبکه هوشمند پیشنهاد گردید. روش پیشنهادی از چهار دسته‌بند پایه SVM، شبکه عصبی MLP، درخت تصمیم و K-NN برای دسته‌بندی داده‌های کنتور هوشمند استفاده می‌کند و نتایج به دست آمده از طریق الگوریتم تقویتی Adaboost بهبود می‌یابد. نتایج به دست آمده نشان می‌دهد که این روش دارای دقت و عملکرد خوبی نسب به سایر روش‌ها است. هم چنین روش ارائه شده از نظر میزان نیاز به نمونه‌برداری نسبت به روش‌های دیگر دارای مقدار کم‌تری است که این امر حریم خصوصی مشتریان را حفظ می‌کند. در آینده می‌توان از الگوریتم ژنتیک برای بهینه‌سازی وزن دسته‌بندها استفاده کرد. هم چنین می‌توان ویژگی‌هایی از قبیل فصل، دما، تعداد افراد خانواده، تعداد لوازم خانگی، میزان برق دریافتی و برق گزارش شده را به سیستم افزود تا دقت تشخیص بهبود یابد.

مراجع

[1] H. Farhangi, "The path of the smart grid," *Power and Energy Magazine, IEEE*, vol. 8, pp. 18-28, 2010.

[2] A. Pyasi and V. Verma, "Improvement in electricity distribution efficiency to mitigate pollution IEEE ISEE (May 2008)," in *Electronics and the Environment, 2008. ISEE 2008. IEEE International Symposium on*, 2008, pp. 1-1.

[3] B. Krebs, "FBI: Smart meter hacks likely to spread," *Krebs on Security. Available online: http://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/(accessed on 25 April 2012)*, 2012.

[4] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. S. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Science and Technology*, vol. 19, pp. 105-120, 2014.

[5] P. Jokar, N. Arianpoo, and V. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns," *Smart Grid, IEEE Transactions on*, vol. 7, pp. 216-226, 2016.

[7] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *Power Delivery, IEEE Transactions on*, vol. 25, pp. 1162-1171, 2010.

[8] J. Nagi, K. Yap, S. Tiong, S. Ahmed, and A. Mohammad, "Detection of abnormalities and electricity theft using genetic support vector machines," in *TENCON 2008-2008 IEEE Region 10 Conference*, 2008, pp. 1-6.

[9] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system," *Power Delivery, IEEE Transactions on*, vol. 26, pp. 1284-1285, 2011.

[10] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision Tree and SVM-based Data Analytics for Theft Detection in Smart Grid".

[11] A. Cardenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, "A game theory model for electricity theft detection and privacy-aware control in AMI systems," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, 2012, pp. 1830-1837

[12] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*: Elsevier, 2011

[13] U. Greveler, P. Glösekötterz, B. Justusy, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, 2012, p. 1.

[14] S. McLaughlin, B. Holbert, S. Zonouz, and R. Berthier, "AMIDS: A multi-sensor energy theft detection framework for advanced metering infrastructures," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, 2012, pp. 354-359.