

کاربرد داده‌کاوی در شناسایی مشترکین با مصارف غیر مجاز آب (مطالعه موردی شرکت آب و فاضلاب قم)

قاسم امینی^{۱*}، حسن فرمانی انتظام^۲، علی جان صادق‌پور^۳، عباس داودآبادی^۴

۱- کارشناس ارشد آمار شرکت آب و فاضلاب شهری استان قم، ghasem2094.amini@gmail.com

۲- دانشجوی دکتری دانشکده مهندسی صنایع، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران،
has.farmanientezam.eng@iauctb.ac.ir

۳- رئیس هیات مدیره و مدیر عامل شرکت آب و فاضلاب شهری استان قم، modiramel@abfa-qom.ir

۴- رئیس اداره آمار و تحلیل اطلاعات شرکت آب و فاضلاب شهری استان قم، adf605@yahoo.com

چکیده

یکی از مصادیق عینی در کاهش مصرف آب بدون درآمد ارائه راه‌کارهای قانونی، عملی و برخورد با برداشته‌های غیرمجاز از شبکه توزیع آب است. انشعابات غیرمجازی که با برداشت آب خارج از مجرای قانونی منجر به عدم اندازه‌گیری مصارف، افزایش هدر رفت آب، کاهش فشار، افزایش احتمال ورود آلاینده‌ها به شبکه، عدم استفاده از اتصالات استاندارد و در نتیجه، نشت آب و ایجاد فرورفتگی‌ها و تخریب اماکن پیرامون شبکه می‌گردد. این مقاله به بررسی و ارائه مدلی جهت تشخیص مصارف غیرمجاز و تخصیص مشترکین به کلاس‌های ایجاد شده می‌پردازد. داده‌های مورد استفاده شامل چهارده دوره میانگین مصرف ماهانه ۱۱۷۰۱۷ مشترک خانگی شهر مقدس قم در سال‌های ۹۶-۹۵ است. به منظور تحلیل داده‌ها از دو روش داده‌کاوی، خوشه‌بندی دومرحله‌ای و شبکه عصبی استفاده شده است. با استفاده از خوشه‌بندی، مشترکین در سه خوشه دسته‌بندی و خوشه شماره سه با تعداد ۳۵۸۷ مشترک با رفتار غیر طبیعی به عنوان کاندیدی برای مصارف غیر مجاز شناسایی شد. در نهایت با استفاده از شبکه عصبی به دسته‌بندی مشترکین خانگی در خوشه‌های حاصل پرداخته شد. درصد پیش‌بینی صحیح بالای ۹۲٪ در هر یک از نمونه‌های آموزش، آزمایش و صحت‌آزمایی برای خوشه سوم حاکی از دقت بالای دسته‌بندی توسط این روش دارد. بنابراین، می‌توان از این روش جهت شناسایی مصارف مشکوک استفاده کرد.

واژه‌های کلیدی: مصرف غیرمجاز، داده‌کاوی، خوشه‌بندی، شبکه عصبی، آب بدون درآمد

۱- مقدمه

امروزه استفاده از انشعابات غیر مجاز آب به صورت آشکار و پنهان یکی از پدیده‌هایی است که مسائل و مشکلات متعددی را برای صنعت آب و فاضلاب به وجود آورده است و خسارات مستقیم و غیرمستقیم ناشی از آن باعث گردیده که طی سالهای اخیر شرکت‌های آب و فاضلاب نگاه ویژه‌ای به این گونه انشعابات داشته و برخورد با عاملین را با انسجام بیشتری مورد پیگیری قرار داده و با اجرای طرح‌های مقابله با انشعابات غیر مجاز آب نقش موثری در کاهش این پدیده ناهنجار داشته باشند.

با توجه به شرایط کنونی بحران آب و افزایش تقاضا برای این ماده حیاتی و کالای ارزشمند، نگاه فراگیر به آب و مدیریت آن از اهمیت ویژه‌ای برخوردار است و از این رو شناسایی مصارف غیرمجاز به عنوان موضوع مهم در تلفات غیر فنی شبکه آب یکی از گام‌های اساسی در توسعه خدمات رسانی و افزایش بهره‌وری از منابع آب به حساب می‌آید که بایستی ضرورتاً با آن توجه خاص نمود.

مشترکین با مصارف غیر مجاز را می‌توان به دو دسته تقسیم نمود. دسته اول مشترکین هستند که کنترل آب دارند ولی بنا بر دلایلی اقدام به دستکاری کنترل می‌کنند. دسته دوم، فاقد اشتراک بوده و کنترل آب ندارند و مستقیم از

شبکه توزیع استفاده می کنند که باعث اختلال در مدیریت توزیع، تخریب شبکه آبرسانی، احتمالاً ایجاد آلودگی و نشت آن به شبکه در هنگام نصب انشعاب و هدر رفت آب به علت استفاده از اتصالات و لوله های غیر استاندارد، افت فشار و در نتیجه نارضایتی مردم را به همراه خواهد داشت.

از جمله عوامل و دلایل فزونی انشعابات غیرمجاز می توان به عوامل فرهنگی، عوامل اداری مرتبط با شرکت های آب و فاضلاب، شهرداری، نهادهای نظارتی (از قبیل طولانی شدن پروسه اداری روند خرید انشعاب، نبود برخی زیرساختهای شبکه در محل) و عوامل اقتصادی اشاره کرد.

افزایش فعالیت های فرهنگی در خصوص آگاهی از عواقب برداشت های غیرمجاز، ایجاد هماهنگی بیشتر بین شرکت آبفا و شهرداری، کوتاه تر شدن مراحل اداری، افزایش آگاهی دفاتر پیشخوان دولت جهت راهنمایی مشترکین و تسریع و تسهیل در روند خرید انشعاب و ارائه هرگونه خدمات به مشترکین، تقسط بیشتر مبلغ حق انشعاب در مناطق محروم، تشکیل واحد منسجم و متمرکز جهت شناسایی و برخورد با انشعابات غیرمجاز، می توانند به عنوان راه کارهایی جهت کاهش انشعابات غیر مجاز در نظر گرفته شوند.

امروزه با پیشرفت تکنولوژی استفاده وسیعی از داده کاوی در تشخیص کلاهبرداری و کشف جرم می شود. تحقیقات متنوعی در این زمینه به ویژه در حوزه مالی یافت می شود. مانند استفاده (2018) navanshu and saad از تکنیک های ماشین بردار پشتیبان، درخت تصمیم، رگرسیون لجستیک در تشخیص تقلب کارتهای اعتباری، تحقیقات (2018) monika and amarpreet و (2018) ernest et al و (2013) anita and ravindra که به بررسی و مقایسه تکنیک های مختلف داده کاوی در تشخیص تقلب می پردازند. در حوزه صنعت آب و برق نیز می توان به مطالعات ذیل اشاره کرد. مدلسازی تخلف در مصرف آب توسط داده کاوی توسط (2012) hashem and humaid، بررسی روش های تشخیص مصرف غیر مجاز برق بر اساس داده های اندازه گیری در ساختار شبکه هوشمند توسط نائینی و قاینی (۱۳۹۶)، ارائه راهکار دو مرحله ای جهت شناسایی الگوی مصرف برق توسط آموزگار (۱۳۹۵)، شناسایی الگوی مصرف انرژی الکتریکی با داده کاوی توسط کجوری و فریدونیان (۱۳۹۴)، تشخیص کلاهبرداری در سازمان های خدمات رسانی با داده کاوی توسط مینایی و دیگران (۱۳۹۰).

مطالعات انجام شده نشان می دهد در داخل کشور تحقیق بسیار کمی در خصوص شناسایی انشعابات غیرمجاز دستکاری شده با رویکرد داده کاوی در شرکت های آب و فاضلاب وجود دارد. این شرکت ها نیز معمولاً این مهم را به صورت سنتی توسط مامورین قرائت (بواسطه تغییرات ظاهری در کنتور) انجام می دهند. لذا تحقیق حاضر از این لحاظ جنبه نوآوری داشته و هدف آن، استفاده از تکنیک های داده کاوی جهت تشخیص مصارف مشکوک به دستکاری در بین مشترکین خانگی با استفاده از سابقه مصرفی آنها می باشد.

ذکر این نکته نیز حائز اهمیت است که امکان در نظر گرفتن همه وضعیت های قرائت مانند در بسته، وجود مانع، کنتور خراب و تخلیه با توجه به تخمین میانگین مصارف آنها وجود ندارد، لذا تنها مشترکینی در نظر گرفته شده اند که در هر دوره قرائت، کنتور آنها سالم و مامور موفق به قرائت کنتور شده است.

۲- مواد و روش ها

۲-۱ خوشه بندی

استخراج و شناسایی الگو از بین حجم انبوه داده ها یک مسئله غیر نظارتی است که با استفاده از روشهای خوشه بندی قابل اجرا است. در این راستا راه کارهای متعددی ارائه شده است که با استفاده از الگوریتم های خوشه بندی، داده های

مشابه را در یک خوشه قرار داده و الگویی را به عنوان نماینده ارائه میکنند. در واقع این نماینده بیانگر رفتار داده های آن خوشه است (آموزگار، ۱۳۹۵). این داده های مشابه بیشترین تشابه را با یکدیگر و کمترین تشابه را با اعضاء خوشه های دیگر داشته باشند. برای خوشه بندی روش های مختلفی از قبیل کا میانگین، دو مرحله ای و سلسله مراتبی وجود دارد. روش کا میانگین، ساده ترین روش خوشه بندی و درداده های کمی کاربرد دارد. روش دو مرحله ای در داده های بزرگ و استفاده همزمان داده های کمی و کیفی کاربرد دارد. روش دو مرحله ای بر خلاف روش های دیگر از تعداد خوشه ها توسط داده ها تعیین می شوند. در حالی که در روش کا میانگین تعداد خوشه ها با استفاده از آزمون و خطا تعیین می شود. روش خوشه بندی دو مرحله ای جهت خوشه بندی از معیارهایی از قبیل AIC و BIC و مقیاس های فاصله اقلیدسی و لگاریتم درستی استفاده می کند.

با توجه به حضور متغیر کمی و متغیر کیفی منطقه در این مطالعه روش مورد استفاده خوشه بندی دو مرحله ای است. این روش در دو گام کار خوشه بندی را انجام می دهد در گام اول همه داده ها یک مرور کلی می شوند و داده های ورودی خام به مجموعه ای از زیر خوشه های قابل مدیریت تقسیم می شوند. گام دوم با استفاده از یک روش خوشه بندی سلسله مراتبی بطور مداوم زیر خوشه ها را برای رسیدن به خوشه های بزرگتر با هم ترکیب می کند بدون اینکه نیاز باشد که جزئیات همه داده ها مجددا مرور کند. این روش می تواند فیلهایی با نوع ترکیبی و همچنین مجموع داده های زیاد را به خوبی پوشش دهد.

۲-۲ شبکه عصبی مصنوعی

شبکه های عصبی مصنوعی یکی از پویاترین حوزه های تحقیق در دوران معاصر می باشد که توجه افراد بسیاری از رشته های گوناگون علمی را به خود جلب کرده است. یکی از بزرگترین مزیت های شبکه عصبی، انعطاف پذیری آن ها برای پیش بینی انواع مدل های غیرخطی است.

شبکه های عصبی یک تقریب زنده جهانی هستند که می توانند هر نوع تابعی را با دقت مورد نظر تقریب بزنند، بدون اینکه نیاز به هیچ گونه پیش فرضی در مورد شکل مدل داشته باشد و همین ویژگی یکی از مزیت های بارز شبکه های عصبی نسبت به مدل های غیرخطی دیگر است.

در اواسط دهه ۱۹۸۰ الگوریتم های مربوط به شبکه عصبی مصنوعی در حل مسائل واقعی به کار گرفته شدند و ابزار ریاضی هستند که با تقلید از سیستم عصبی بیولوژیک ساخته شده اند (Fulop et al, 1998)، قدرت انعطاف و تصحیح پذیری بالایی در انطباق خود با داده های موجود دارند (Ranjithan et al, 1995) به گونه ای که قادرند به سازماندهی مجهز شده، نظم و هماهنگی موجود در داخل این داده ها را پیدا کنند (منهاج، ۱۳۸۴) و بر اساس بردارهای ورودی، رخداد و بزرگی یک پدیده را پیش بینی نمایند. در واقع شبکه عصبی مصنوعی مجموعه ای از نرون های به هم متصل در لایه های مختلف است. این نرون ها اطلاعاتی را برای یکدیگر ارسال کرده و واحدهای ساده پردازش اطلاعات بوده و تعداد زیادی از آنها یک شبکه عصبی مصنوعی را می سازد (آرام و کهنه شهری، ۱۳۹۱).

در ساخت یک مدل بر مبنای شبکه عصبی اولین کار انتخاب نوع شبکه است. پس از آن پارامترهای ورودی که در خروجی تاثیرگذار هستند، انتخاب می گردند. سپس معماری شبکه یعنی، تعداد لایه ها و تعداد گره ها در هر لایه و چگونگی اتصال آنها، و نوع توابع محرکه مورد استفاده برای نرون ها و نهایتاً پارامترهای موثر در آموزش شبکه تعیین می گردند.

بعد از تعیین نوع شبکه و معماری آن یک سری آمار ورودی و خروجی به شبکه داده می شود، شبکه بر اساس این اطلاعات آموزش می یابد. این مرحله، مرحله یادگیری یا آموزش شبکه نامیده می شود. در مرحله آموزش با توجه به نوع الگوریتم آموزشی، وزن های اتصالات شبکه تغییر می یابند. مناسب ترین مدل، با سعی و خطا به صورت تغییر

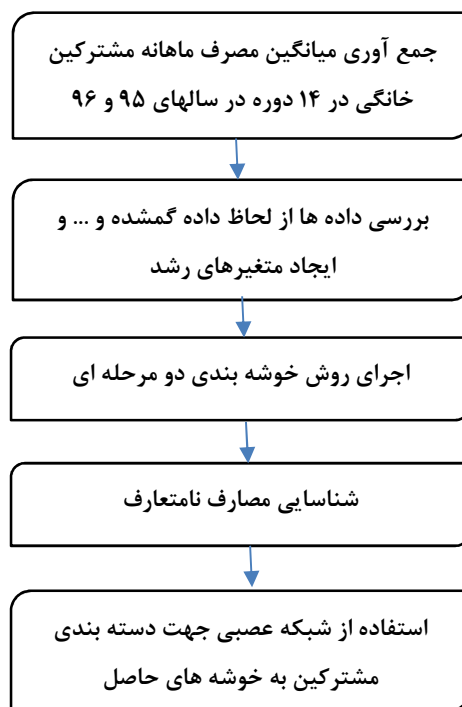
توابع محرکه، تغییر تعداد لایه ها، تغییر تعداد گره ها در لایه های پنهان، به دست می آید. بعد از آموزش، شبکه با استفاده نمونه آزمایش مورد تجدید نظر قرار می گیرد و صحت عملکرد شبکه امتحان می گردد تا بدین طریق خطای حاصل از برآورد کاهش یابد. به عبارتی نمونه آزمایشی با کمک به نمونه آموزش خطای حاصل را کاهش می دهد. در ادامه شبکه با استفاده از نمونه صحت آزمایی که در محاسبات منظور نمی گردد و جهت ارزیابی دقت مدل مورد استفاده قرار میگیرد.

در این مقاله به منظور دستیابی به اهداف مورد نظر در اینجا دسته بندی، از شبکه عصبی چند لایه پیشخور (MLP) استفاده گردید و سعی گردید تا بهترین و کاراترین شبکه با تعیین مقدار خطای آن انتخاب و مورد استفاده قرار گیرد. این شبکه ها دارای سه لایه به نامهای لایه ورودی، لایه میانی (پنهان) و لایه خروجی می باشند. نخستین لایه را، لایه ورودی و آخرین لایه را، لایه خروجی می نامند. به لایه میانی نیز لایه پنهان گفته می شود.

۲-۳ داده های پژوهش

جامعه مورد بررسی در این پژوهش مشترکین خانگی شرکت آب و فاضلاب شهر قم می باشند. جمعیت این شهر طبق سرشماری سال ۱۳۹۵ بالغ بر یک میلیون و دویست هزار نفر می باشد و مشترکین خانگی با اختصاص ۷۶ درصد مصرف آب این شهر بیشترین سهم مصرف را به خود اختصاص داده اند.

داده های مورد بررسی مربوط به میانگین مصرف ماهانه هفت دوره قرائت مشترکین خانگی شهر قم در سالهای ۹۵ و ۹۶ می باشد. با بررسی اطلاعات پایگاه داده در میان تمام مشترکین خانگی تا پایان سال ۱۳۹۶، تعداد ۱۱۷۰۱۷ مشترک که در ۱۴ دوره قرائت سالهای ۹۵ و ۹۶ دارای وضعیت قرائت سالم، مصرف غیر صفر، عدم تصحیح قرارداد (از قبیل تغییر کاربری، واحد شمار، تعویض کنتور) بوده اند، به عنوان نمونه انتخاب و با استفاده از روش های داده کاوی خوشه بندی دو مرحله ای و شبکه عصبی به شناسایی الگوی مصارف غیرمجاز پرداخته شد. در این پژوهش از نرم افزار Spss جهت تحلیل داده استفاده شده است. شکل ۱ فلوچارت مراحل کار را نشان می دهد.



شکل ۱: فلوچارت مراحل روش تحقیق

۳- نتایج و بحث

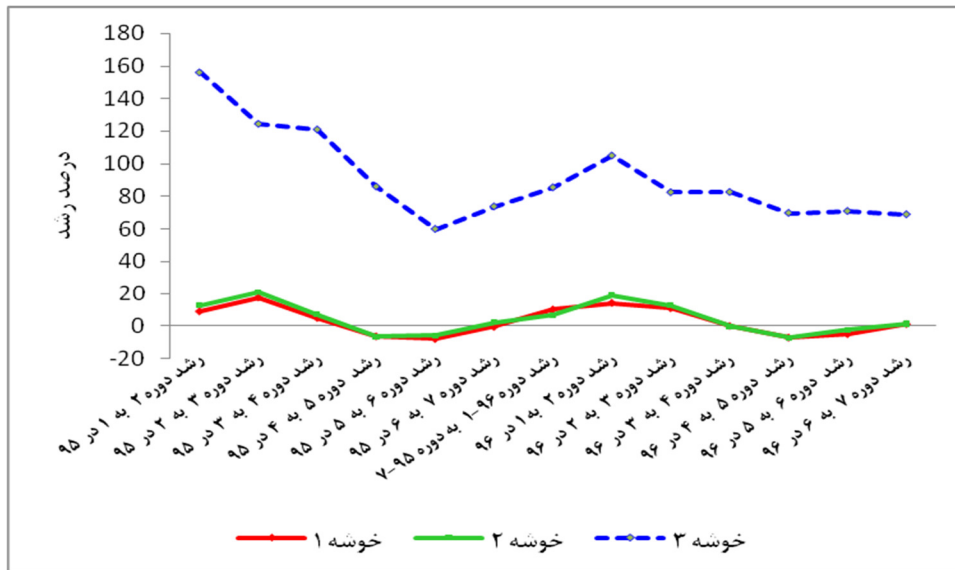
۳-۱- تشخیص مصارف غیر مجاز آب توسط خوشه بندی

در این بخش به منظور شناسایی انشعابات غیر مجاز از تکنیک داده کاوی و روش خوشه بندی دو مرحله ای و شبکه عصبی جهت دسته بندی مشترکین به خوشه به دست آمده استفاده شده است. متغیرهای مورد استفاده در این مرحله شامل ۱۳ متغیر کمی رشد مصرف دوره های متوالی در سالهای ۹۵-۹۶ و همچنین منطقه به عنوان متغیر گسسته می باشند. بعد از اجرای خوشه بندی، تعداد ۳ خوشه بهینه با کیفیت ۰/۹ با استفاده از معیارهای فاصله درستی ماکزیمم و بیزین بصورت خودکار تعیین شد.

جدول ۱ و شکل ۲ خوشه ها به همراه نمایندگان رشد هر دوره را نشان میدهند. خوشه ۱ و ۲ به ترتیب شامل تمام مشترکین مناطق ۱ و ۲ شهر قم و خوشه ۳ نیز شامل ۶۵٪ مشترکین منطقه ۱ و ۳۵٪ مشترکین منطقه ۲ می باشند. تغییرات رشد در خوشه های ۱ و ۲ با روندی سینوسی شبیه هم و متفاوت از خوشه ۳ با روند نزولی می باشند. در خوشه ۳ رشد دوره های متوالی بسیار زیاد است و از طرفی در دوره های ۵ به ۴ و دوره ۶ به ۵ که معمولا مصرف کمتر می شود، انتظار رشد منفی داریم (همانگونه که در دو خوشه دیگر چنین است) در حالی که چنین رفتاری در خوشه ۳ برعکس می باشد و شاهد افزایش رشد مصرف در این دوره ها هستیم. بنابراین خوشه ۳ با تعداد ۳۵۸۷ اشتراک رفتار نرمالی از خود نشان نداده و می تواند به عنوان کاندید دستکاری و انشعابات غیر مجاز در نظر گرفته شود.

جدول ۱: نتایج خوشه بندی رشد مصرف مشترکین خانگی شهر قم در سالهای ۹۶-۹۵

خوشه ۳	خوشه ۲	خوشه ۱	رشد دوره
۱۵۶	۱۳	۹	رشد مصرف دوره ۲ نسبت به دوره ۱ در سال ۹۵
۱۲۵	۲۱	۱۷	رشد مصرف دوره ۳ نسبت به دوره ۲ در سال ۹۵
۱۲۱	۷	۵	رشد مصرف دوره ۴ نسبت به دوره ۳ در سال ۹۵
۸۶	-۷	-۶	رشد مصرف دوره ۵ نسبت به دوره ۴ در سال ۹۵
۶۰	-۶	-۸	رشد مصرف دوره ۶ نسبت به دوره ۵ در سال ۹۵
۷۴	۲	-۰/۴	رشد مصرف دوره ۷ نسبت به دوره ۶ در سال ۹۵
۸۵	۷	۱۰	رشد مصرف دوره ۱ از سال ۹۶ نسبت به دوره ۷ در سال ۹۵
۱۰۵	۱۹	۱۴	رشد مصرف دوره ۲ نسبت به دوره ۱ در سال ۹۶
۸۲	۱۳	۱۱	رشد مصرف دوره ۳ نسبت به دوره ۲ در سال ۹۶
۸۳	-۰/۱۷	۰/۱	رشد مصرف دوره ۴ نسبت به دوره ۳ در سال ۹۶
۷۰	-۷	-۷	رشد مصرف دوره ۵ نسبت به دوره ۴ در سال ۹۶
۷۱	-۲	-۵	رشد مصرف دوره ۶ نسبت به دوره ۵ در سال ۹۶
۶۹	۱	۱	رشد مصرف دوره ۷ نسبت به دوره ۶ در سال ۹۶
۳۵۸۷	۵۶۱۳۸	۵۷۲۹۲	تعداد مشترک
۳	۴۸	۴۹	درصد



شکل ۲: روند رشد مصرفی دوره ها به تفکیک خوشه در کاربری خانگی شهر قم طی سالهای ۹۵-۹۶

۳-۲ دسته بندی مشترکین با استفاده از شبکه عصبی مصنوعی

اغلب از خوشه بندی به عنوان اولین گام فرایند های داده کاوی یاد می شود که قبل از سایر فرایندها برای شناسایی گروهی از رکوردهای مرتبط با هم که بعدا بتوانند نقطه آغاز تحلیل ها باشند بر روی رکوردها اعمال می شود. در این مطالعه با استفاده از داده های رشد مصرف مشترکین خانگی در دوره های متوالی و اجرای روش خوشه بندی دو مرحله ای تعداد ۳ خوشه بهینه به دست آمد که از آن می توان در شناسایی مصارف غیرمجاز استفاده نمود. در ادامه با استفاده از روش شبکه عصبی مصنوعی به مدلسازی عضویت مشترکین به این ۳ خوشه به دست آمده پرداخته می شود. ۷۰٪ داده ها (تعداد ۸۱۱۵۹ اشتراک) به عنوان نمونه آموزش، ۱۵٪ به عنوان نمونه آزمایش (تعداد ۱۷۹۷۹ اشتراک) و ۱۵٪ به عنوان نمونه صحت آزمایی (تعداد ۱۷۸۷۹ اشتراک) در نظر گرفته شد. با توجه به مجموع مربعات خطا و همچنین خطای نسبی مدل های مختلف، در نهایت مدل شبکه عصبی با یک لایه پنهان و ۷ گره در این لایه، تابع محرکه تانژانت هیپربولیک به عنوان تابع لایه پنهان و تابع سافت مکس به عنوان تابع محرکه لایه خروجی به عنوان مناسب ترین مدل برای دسته بندی انتخاب گردید. درصد پیش بینی نادرست هر یک از نمونه آموزش ۰/۴ درصد، نمونه آزمایش ۰/۳ درصد و نمونه صحت آزمایی ۰/۴ درصد به دست آمد. به عبارتی درصد پیش بینی صحیح هر یک از نمونه ها ۹۹/۶ درصد می باشد که نشان از دسته بندی مناسب مشترکین خانگی به خوشه های ۳ گانه است.

با استفاده از این مدل می توان با در اختیار داشتن مصارف جدید، مشترک را در یکی از ۳ خوشه طبقه بندی کرد و با توجه به اینکه خوشه ۳ موارد مشکوک به مصارف غیر مجاز می باشند لذا مشترک جدیدی که به این خوشه تعلق گیرد، می تواند به عنوان موردی مشکوک از مصارف غیرمجاز مورد بررسی قرار گیرد. جدول ۲ نتایج دسته بندی و درصد پیش بینی صحیح را نشان می دهد.

جدول ۲: نتایج دسته بندی توسط شبکه عصبی بر اساس شاخص رشد مصرف

نمونه	خوشه واقعی	خوشه پیش بینی		
		خوشه ۱	خوشه ۲	خوشه ۳
نمونه آموزش	خوشه ۱	۳۹۷۲۸	۰	۴۸
	خوشه ۲	۰	۳۸۸۰۶	۹۹
	خوشه ۳	۷۰	۱۲۸	۲۲۸۰
نمونه آزمایش	خوشه ۱	۸۷۸۲	۰	۵
	خوشه ۲	۰	۸۶۱۴	۱۸
	خوشه ۳	۸	۲۹	۵۲۳
نمونه صحت آزمایی	خوشه ۱	۸۷۱۳	۰	۱۶
	خوشه ۲	۰	۸۵۷۹	۲۲
	خوشه ۳	۱۲	۲۴	۵۱۳

۴- نتیجه گیری

این مقاله با استفاده از روش های داده کاوی خوشه بندی دو مرحله ای و شبکه عصبی به شناسایی مصارف نامتعارف و مشکوک به دستکاری کنتور آب پرداخته است. بدین منظور متوسط مصرف ماهانه ۱۱۷۰۱۷ مشترک خانگی شهر قم طی چهارده دوره در سالهای ۹۶-۹۵ استخراج و روش های اشاره شده بر روی تغییرات مصرف دوره ها اجرا شد. نتایج نشان می دهد ۹۷ درصد مشترکینی که در تمام دوره ها دارای تغییرات عادی بودند در دو خوشه و ۳ درصد آنها (۳۵۸۷ مشترک) با تغییرات نامنظم در خوشه ای جداگانه قرار گرفتند. این خوشه می تواند در پیمایش ها به عنوان کاندیدی برای بررسی موارد مشکوک در نظر گرفته شود. دقت بالای نتایج شبکه عصبی در دسته بندی مشترکین به خوشه ها، بیانگر کارآمد بودن و استفاده از این روش در شناسایی مصارف نامتعارف دارد.

۵- مراجع

- آرام، ع. و عاقلی کهنه شهری، ل. (۱۳۹۱)، "ارائه یک مدل ترکیبی برای پیش بینی تقاضای روزانه آب شهری"، *فصلنامه اقتصاد مقداری (بررسی های اقتصادی سابق)*، ۲۹(۲)، ۱-۱۷.
- آموزگار، م. (۱۳۹۵)، "ارائه راهکار دو مرحله ای جهت شناسایی الگوی مصرف برق"، *نشریه علمی پژوهشی کیفیت و بهره وری صنعت برق ایران*، ۵ (۹)، ۴۸-۵۷.
- کجوری نفت چالی، م. و فریدونیان، ع. (۱۳۹۴)، "شناسایی الگوی مصرف انرژی الکتریکی با داده کاوی"، *سی / امین کنفرانس بین المللی برق، تهران*.
- کسانیان نائینی، ا. و قاینی، م. (۱۳۹۶)، "بررسی روش های تشخیص مصرف غیر مجاز برق بر اساس داده های اندازه گیری در ساختار شبکه هوشمند"، *سی و دومین امین کنفرانس بین المللی برق، تهران*.
- منهاج، م. (۱۳۸۴)، *مبانی شبکه های عصبی (هوش محاسباتی)*. جلد ۱، چاپ سوم، مرکز نشر دانشگاه صنعتی امیر کبیر.
- مینایی، ب.، دیانت، ر.، هانی، ح. و سبحانی نیا، م. (۱۳۹۰). "تشخیص کلاهبرداری در سازمان های خدمات رسانی با داده کاوی". پایان نامه کارشناسی ارشد رشته فناوری اطلاعات، دانشگاه قم



Anita, B. D., and Ravindra, D., (2013), " Data mining techniques for fraud detection", Journal of Computer Science and Information Technologies, 4(1), pp 1-4.

Ernest, O. N., Chukwuedozie, N. E., and Inyama, C. H., (2018), " Review of Fraud Detection and Churn Behavior Modeling Techniques", Journal of Engineering Technology Open Access, 2(1).

Fulop, I. A., Jozsa, J., and Karamer. T., (1998), "A neural network application in estimating wind induced shallow lake motion", Journal of Hydro informatics, 98(2), pp 753-757.

Hashem, E. and Humaid, S., (2012), "A data mining based fraud detection model for water consumption billing system in MOG", MSc. Thesis, Department of Computer Science, Islamic University of Gaza, Gaza.

Monika, C., Amarpreet, K.,(2018), "A Comparative Study of Classification Techniques for Fraud Detection", Journal on Future Revolution in Computer Science & Communication Engineering, 4(5), pp 19-23.

Navanshu, K., Saad, Y.S., (2018)," Credit card fraud detection using machine learning modeles and collating machine learning models", Journal of Pure and Applied Mathematics, 118(20), pp 825-838.

Ranjithan, J., Eheart, J., and Garrett, J. H., (1995), Application of neural network in ground water remediation under condition of uncertainty, New Uncertainty Conceptsion Hydrology and Water Resources.

Application of data mining to identify subscribers with unauthorized use of water (Case Study of Qom Water and Wastewater Company)

**Ghasem amini*¹, Hasan Faramani Entezam², Alijan Sadeghpour³,
Abbas Davood Abadi⁴**

**1- MSc. of Statistics of Qom Province Water and Wastewater company.
Ghasem2094.amini@gmail.com**

**Engineering, Central Tehran Department of Industrial .Of .PHD student
Branch, Islamic Azad University, Tehran has.farmanientezam.eng@iauctb.ac.ir**

**Chairman of the Board And managing director of Qom Province wwc
modiramel@abfa-qom.ir**

**Chief of Statistics and Analysis Department of Qom Province wwc
adf605@yahoo.com**

Abstract

One of the concrete examples in reducing consumption Non-revenue water provides legal, practical, and non-permissible solutions to the water distribution network. Unauthorized water connections, which are caused by the removal of water outside the legal duct, lead to a lack of cost measurement, increased water losses, reduced pressure, increased likelihood of entering pollutants into the network, non-use of standard connections, and consequently water leakage and damming and destruction. Places around the network. This article examines and provides a model for detecting unauthorized charges and assigning subscribers to created classes. The data used include fourteen periods of the average monthly consumption of 117017 households in the sacred city of Qom in the years 2017-2018. For data analysis, two methods of data mining, two-stage clustering and neural network have been used. Using clustering, subscribers were identified in three cluster cluster and cluster number three with 3587 subscribers with abnormal behavior as a candidate for unauthorized use. Finally, by using the neural network, the household subscribers were ranked in the resulting clusters. The correct prediction percentage of up to 92% in each of the training, testing, and verification samples for the third cluster indicates a high degree of classification by this method. Therefore, this method can be used to identify unauthorized costs.

Keywords: unauthorized use, data mining, clustering, neural networks, non-revenue water