

مدیریت دانش در شناسایی الگوی مصرف آب با کمک داده کاوی (مطالعه موردی: شهر فسا)

عباس دهقان بهرغانی^۱، دکتر محمد رضا فیلی زاده^۲

^۱ مدیر دفتر تجهیز منابع مالی شرکت آب و فاضلاب استان فارس، Dehghan55@gmail.com

^۲ استادیار دانشگاه آزاد اسلامی شیراز، Feylizadeh_mr@yahoo.com

چکیده:

پیاده سازی اصلاح الگوی مصرف آب یعنی نهادینه کردن روش صحیح استفاده از منابع آب شرب کشور که این امر سبب ارتقای شاخص‌های زندگی و کاهش هزینه‌ها در بخش آب می‌شود. اولین قدم در این زمینه شناسایی رفتار مصرفی و الگوی مصرف آب می‌باشد تا مشخص شود علایق مشترکین در کاربری‌های مختلف برای مصرف آب به چه صورت می‌باشد، بعد از آن زمینه برای ارائه برنامه‌ها و راهکارهای اصلاح الگوی مصرف تسهیل تر می‌گردد. مدیریت دانش یکی از کارآمدترین روش‌هایی است که بوسیله آن الگوهای مفید نهفته در داده‌ها شناخته می‌شود و اطلاعاتی را در اختیار تحلیل گران قرار می‌دهد تا بر اساس آن‌ها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شود، یکی از ابزارهای مهم در پیاده سازی مدیریت دانش، داده کاوی است. در این مقاله با کمک الگوریتم **K-Medoids**، الگوی مصرف مشترکین آب در تمام کاربری‌ها شناسایی می‌شود. نتایج بدست آمده حاکی از آن است که پیاده‌سازی تکنیک‌های مدیریت مصرف در کاربری‌های مختلف با توجه به الگوی هر کاربری قابل برنامه ریزی و اجرا خواهد بود و تحقق این امر تنها در شناسایی الگوی مصرف کاربری‌ها خواهد بود. همچنین، با کمک نتایج و اطلاعات کسب شده، پیشنهادات کاربردی برای شرکت آبفا فارس در راستای بالا بردن کارایی خود در انجام امور مربوط به کنترل و نظارت بر مصرف مشترکین آب می‌باشد

کلمات کلیدی: تحلیل رفتار مشترکین، داده کاوی، مدیریت دانش.

۱- مقدمه

آب و رقابت جهت سیطره بر منابع محدود آن، یکی از مهمترین حوزه‌های چالش برانگیز هزاره سوم خواهد بود. رشد جمعیت، توسعه صنعتی و اجتماعی و تغییرات آب و هوایی هر یک از سوی منابع محدود آب سالم را تحت فشار قرار داده‌اند. اگرچه آب سالم در طول تاریخ بشر هرگز کالای کم ارزشی نبوده، اما تصور خطرناک فراوان انگاری آب، موجب گردیده است که بسیاری از مردم جهان بر اثر عدم توجه کافی به مدیریت صحیح تولید و مصرف به مرزهای بحران آب برسند. بحران آب زودتر از اغلب کشورهای جهان دامنگیر کشور خشک ایران شده است. کم آبی چند ساله اخیر اگر چه در نگاه بسیاری امری گذرا تلقی می‌شود، ولی در حقیقت می‌توان آن را پیش قراول بحران

گسترده آب دانست که در صورت عدم برنامه‌ریزی منسجم بدون شک لطمات جبران ناپذیری را بر پیکره آسیب پذیر اقتصاد کشور وارد خواهد ساخت. در این راستا تغییر در نگرش مقوله مدیریت مصرف می‌تواند به عنوان یک راهکار اساسی برای این چالش در نظر گرفته شود. تغییر در نگرش مقوله مدیریت مصرف می‌تواند به عنوان یک راهکار اساسی برای این چالش در نظر گرفته شود به عبارتی به دنبال شناسای مدل‌های مصرف مشترکین و تحلیل مصارف آن‌ها هستیم و سپس با مشخص نمودن رابطه میان اطلاعات بدست آمده به سمت مشخص نمودن یک الگوی مصرف در جمعیت انتخاب شده مشترکین آب می‌باشیم [1]. در این راستا یکی از تکنیک‌های رایج برای استخراج اطلاعات و کشف الگوی مصرف کاربران استفاده از تکنیک مدیریت دانش است. چراکه با افزایش اطلاعات الکترونیکی، نیاز به ابزارها و تکنیک‌هایی برای بیرون کشیدن، آنالیز و ارتباط برقرار کردن بین انبوهی از اطلاعات که بسیاری از آنها دارای روابط موقتی و وابسته زمانی و مکانی هستند لازم می‌باشد. تمرکز بر این مساله منجر به خروج الگوها و ارتباطات معنی دار در مجموعه داده‌ها می‌گردد. کشف الگوهای پنهان در داده‌ها دشوار هستند زیرا این الگوها در مقیاس کوچک هستند و درک آنها در محیطشان که بصورت روزانه کاربرد دارند مشکل می‌باشد و نیاز به تجزیه و تحلیل‌های گسترده‌ای دارد. یکی از مرحله‌های ضروری از فرآیند کشف دانش از بانک‌های اطلاعاتی¹ (KDD)، داده کاوی است. داده کاوی فرآیندی است که از ابزارهای تحلیلی گوناگونی برای کشف الگوها و روابط بین داده‌ها استفاده می‌کند که ممکن است برای اعتبار بخشیدن به پیش بینی استفاده شود [2].

داده کاوی فرآیندی است، کاملاً علمی جهت کشف دانش‌های پنهان و روابط ناشناخته بین داده‌ها، که در آغاز دهه ۹۰ پا به عرصه ظهور گذاشته و با نگرشی نو، به مسئله استخراج داده‌ها می‌پردازد [3]. در اصل سال ۱۹۸۹ و ۱۹۹۱ که کارگاه‌های کشف دانش از پایگاه داده‌ها، توسط یانتسکی و همکاران برگزار گردید، به نوعی آغاز عملیات این شاخه از علم است [4]. فرآیند داده کاوی به عنوان یک سیستم توسعه‌ای، به سرعت در حال رشد است و در حوزه‌های گوناگونی بکار گرفته می‌شود. از جمله در حوزه سازمان‌های مربوط به آب و موارد مشابه با آن کارهای متنوعی با استفاده از ابزارهای داده کاوی جهت مطالعه مصرف انجام شده است. به عنوان مثال بابویک^۲ و همکاران [۵] از الگوریتم ژنتیک و مدیریت ریسک، جهت تعیین ریسک ترکیدن لوله‌های آب استفاده نمودند. به این صورت که از داده‌های موجود در زمینه ترکیدگی لوله‌ها در زمان‌های گذشته استفاده کرده تا یک مدل ریسک را بعنوان تابعی از مشخصات لوله‌های ترکیده مانند عمر، قطر و جنس ایجاد کنند. گیوس^۳ و همکاران [6] نیز، در کشور انگلستان با استفاده از الگوریتم جدید رگرسیون تکاملی چند جمله‌ای^۴ برای پیش بینی میزان ترکیدگی شبکه‌های توزیع آب استفاده نموده، که در نهایت این کار زمینه مدیریت و نوسازی سیستم آب را فراهم نمود. در مقاله‌ای دیگر که توسط دیاز^۵ و همکاران [7] ارائه شده، بیشتر بحث بر فعالیت‌های آماده سازی داده‌ها برای فعالیت‌های داده کاوی تاکید داشته است. در این مقاله فعالیت‌های تجمیع داده‌ها از پایگاه داده‌های مختلف تا داده‌های گمشده و هم چنین داده‌های پرت، مورد توجه قرار گرفته شده و از دو ابزار شبکه عصبی- مصنوعی و درخت تصمیم برای پیدا کردن داده‌های گمشده و پرت، بر روی داده‌های شبکه توزیع آب، استفاده شده است. از دیگر فعالیت‌ها انجام شده در این مقاله، ارائه و نمایش مناسب داده نیز مورد توجه و بررسی قرار گرفته است. بویل^۶ و همکاران [8] نیز بر اساس قبوض مشترکین سازمان‌های ارائه کننده خدمات عمومی، به شناسایی و دسته بندی بهتر خصوصیات مشترکین، به کمک الگوریتم خوشه‌بندی پرداختند. آن‌ها هدف از تحقیق خود را کمک به تصمیم گیری در زمینه برنامه ریزی و روش ارتباطات با مشترکین در زمینه فروش خدمات، بیان کردند.

از آنجا که کنترل حوادث شبکه یکی از ارکان‌های مهم کاهش هدر رفت آب در شبکه‌های توزیع می‌باشد، اولیوریا^۷ و همکاران [9] با استفاده از روش آمارهای کاوشی^۸ نسبت به شناسایی الگوی لوله ترکیدگی بر مبنای حوادث واقعی اتفاق افتاده در شبکه‌های فیزیکی اقدام نموده است. دوتا^۹ [10] نیز با توجه به اهمیت موضوع آب شرب در جهان با بکارگیری برنامه‌های داده کاوی در بخش آب و الگوریتم‌های ANN^{۱۰} و فازی^{۱۱} و همچنین، داده‌های موجود در سیستم‌های اطلاعاتی جغرافیایی^{۱۲} به تهیه مدلی برای دسته بندی آلاینده‌های آب

¹ Knowledge Discovery in Database

² Babovic

³ Giustolisi

⁴ Evolutionary Polynomial Regression (EPR)

⁵ Diaz

⁶ Boyle

⁷ Oliveria

⁸ Scan Statistics

⁹ Dutta

¹⁰ Artificial Neural Networks

اقدام نمود. در یک تحقیق دیگر که توسط خان^{۱۳} و همکاران [11] در استرالیا انجام گرفت، ارزیابی عملکرد روش‌های مختلف داده کاوی برای پیش بینی نیازهای آبی مبنای کار قرار گرفت. در این تحقیق ابتدا با کمک شش روش مختلف داده کاوی از جمله: درخت تصمیم، شبکه های عصبی، ماشین‌های بردار و . . . نسبت به تهیه مدل اقدام شد و در نهایت روش درخت تصمیم بهترین عملکرد در پیش بینی نیازهای آبی را نشان داد.

بعلاوه ون^{۱۴} و همکاران [12] با کمک روش‌های خوشه‌بندی و بکارگیری الگوریتم K-Means و همچنین، سری‌های زمانی که از دیگر الگوریتم‌های مهم در تکنیک داده کاوی است، از داده‌های موجود در بانک‌های اطلاعاتی استفاده کرده و اقدام به تهیه یک نرم افزار کاربردی کرد که خروجی آن پیش بینی آینده مقدار مصرف بود. همچنین در کشور چین بر اساس مطالعات صورت گرفته، با استفاده از الگوریتم DBSCAN^{۱۵} و همچنین نمودار همسایگی (KNN)^{۱۶} میزان شکستگی و اتفاقات صورت گرفته در شبکه های آب توسط خوشه‌بندی فضایی مورد تجزیه و تحلیل قرار گرفته و سپس، با استفاده از خوشه‌بندی مبتنی بر چگالی محدوده‌هایی از شبکه تامین آب که در معرض اتفاقات و لوله ترکیدگی هستند شناسایی شد [13].

در ایران نیز، آقا بابایی و همکاران [14] از دانش داده کاوی در راستای شناسایی مشترکین بدهکار در شرکت آب و فاضلاب مشهد (منطقه ۳) استفاده نمودند. در این تحقیق، سعی در تدوین الگوی شناسایی ویژگی‌ها و رفتار مشترکین بد حساب، و هم چنین شناسایی آن مشترکین، به کمک الگوریتم شبکه عصبی، شده است. هم‌چنین با استفاده از درخت تصمیم متغیرهای موثر در بدهکار شدن مشترکین شناسایی شده و افقی مناسب برای کاهش هزینه‌ها در شرکت آب، ایجاد شده است. یکی از مهمترین نتایج این مقاله پی بردن به این موضوع بود که هرچه زمان رسیدن قبض بدست مشتری، بیشتر از ۱۵ روز طول بکشد، با احتمال بیشتری هم پرداخت نشدن قبوض آب را بدنبال دارد.

بنابراین با توجه به آن چه گفته شد، موضع آب و نحوه مصرف آن امروزه از اهمیت بالایی برخوردار است و ارائه هر گونه راهکاری برای کاهش و مصرف و جلوگیری از هدررفت آن چه در تمامی مراحل تولید، توزیع و مصرف بسیار بارز است. در این راستا، در این مقاله تلاش شده است تا با استفاده از تکنیک داده کاوی و الگوریتم خوشه‌بندی به کشف روابط و قوانین موجود در بین داده‌های مصرف مشترکین آب پرداخته و از این طریق به الگوی مصرف مشترکین دست یافت. به این صورت که با بررسی مقدار و انواع مصرف در یک زمان یا دوره زمانی توسط افراد یا گروه‌های خاص، منحنی تقاضای (مصرف) شبکه در دوره‌های مختلف را شناسایی نمود و مشخص کرد که آیا شبکه پاسخگوی مقدار تقاضا است یا خیر؟ همچنین مدل‌های موجود میان داده‌های مصرف مشترکین آب را تعیین، تا در نهایت الگوی مصرف بر اساس خروجی اطلاعات حاصل شده تهیه و از این طریق مشخص نمود که شرکت آب و فاضلاب چه الگوها و مشکلاتی را می‌تواند در جهت کنترل منابع آب شهری و بهینه‌سازی توزیع آب، شناسایی کند تا بتواند در آینده به وسیله اقدامات و تدابیر و تصمیمات لازم مدیریتی به رفع آن‌ها بپردازد. برای طراحی این مدل از اطلاعات مربوط به مشترکین شهر فسا که در استان فارس قرار دارد استفاده شده است.

ساختار ادامه این مقاله بدین صورت است، در بخش دوم به شرح روش مدنظر تحقیق پرداخته شده است و الگوریتم مدنظر در این مقاله توضیح داده شده است؛ در بخش سوم توضیحاتی در رابطه با ناحیه مورد مطالعه و مجموعه داده‌ها آورده شده است؛ بخش چهارم اختصاص به نحوه پیاده‌سازی مساله و استخراج نتایج دارد؛ و در نهایت در بخش پنجم نتیجه‌گیری و پیشنهاداتی برای تحقیقات آتی ذکر شده است.

۲- شرح روش تحقیق

هدف اصلی از طرح این مقاله ارائه روشی برای خوشه‌بندی اطلاعات مصرف مشترکین آب می‌باشد با این منظور که بتوان الگوی مصرف را از آن استخراج نمود. در این مقاله، روش طرح و تحقیق پیمایشی، تحقیق از نوع موردی، زمان گردآوری از نوع مقطعی، ماهیت

¹¹ Fuzzy

¹² Geographic information system (GIS)

¹³ Khan

¹⁴ Wen

¹⁵ Density Based Spatial Clustering of Applications with Noise

¹⁶ K-Nearest Neighbor

داده‌ها نیز از نوع کمی و براساس هدف از نوع اکتشافی است. ابزار گرد آوری داده‌ها برای این هدف مطالعات کتابخانه‌ای شامل مقالات مرتبط، پایگاه اینترنتی، کتاب‌های مرتبط، اسناد و مدارک مربوط به مدیریت مصرف، روزنامه‌ها و نشریات و همچنین، مصاحبه و مشورت با تعدادی از خبرگان، کارشناسان، مدیران و متخصصان مربوطه بوده است. رویه کار نیز بدین صورت بوده که اطلاعات پایش شده با استفاده از نرم‌افزار متلب خوشه‌بندی شده است.

الگوریتم بکار رفته در این مساله الگوریتم K-Medoids است. الگوریتم K-Medoids جزء دسته الگوریتم‌های خوشه‌بندی مبتنی بر افراز داده‌ها است، دلایل انتخاب این الگوریتم این است که از یک طرف داده‌های مورد استفاده در این مقاله کمی هستند و هدف این است که مشخص شود هر داده در چه خوشه‌ای قرار می‌گیرد، بنابراین باید از روش‌های خوشه‌بندی افزاری مانند الگوریتم K-Means، یا K-Medoids، یا Bisecting K-Means و یا Fuzzy C-Means استفاده شود. از طرف دیگر مزیت الگوریتم K-Medoids نسبت به الگوریتم‌های K-Means، Bisecting K-Means و Fuzzy C-Means این است که چنانچه در داده‌های مساله، رکودهای نويز و خارج از محدود وجود داشته باشد، حساسیت این الگوریتم نسبت به سایرین کمتر خواهد بود و میانگین هر خوشه تغییرات کمتری را خواهد داشت، بعلاوه در الگوریتم Fuzzy C-Means حساسیت بالایی نسبت به حدس‌های اولیه وجود دارد که ممکن است این امر باعث شود الگوریتم در مینیمم‌های محلی توقف شود [15]، بنابراین در این مقاله از الگوریتم K-Medoids استفاده شد. در بخش ۲-۱ الگوریتم K-Medoids مورد استفاده در این مقاله شرح داده شده است.

۲-۱- شرح الگوریتم K-Medoids

خوشه‌بندی یک تابع کاوشی نظارت شده داده کاوی به منظور کشف گروه بندی طبیعی درون داده‌هاست. یک خوشه به مجموعه از داده‌ها اطلاق می‌شود که از جهاتی شبیه به هم دیگر هستند. الگوریتم‌های خوشه‌بندی به طور خودکار ویژگی‌های متمایز کننده‌ی زیر گروه‌ها را تعریف می‌کنند و آنها را سازماندهی می‌نمایند و مدل را منحصرأ از روی روابطی که در داده‌ها وجود دارند و همچنین از روی خوشه‌هایی که الگوریتم شناسایی می‌نماید، آموزش می‌دهند. به عبارتی دیگر، خوشه‌بندی در واقع یافتن ساختار در مجموعه‌ای از داده‌هایی است که طبقه بندی نشده‌اند. به بیان دیگر می‌توان گفت که خوشه‌بندی قراردادن داده‌ها در گروه‌هایی است که اعضای هر گروه از زاویه‌ی خاصی به یکدیگر شباهت دارند و با اعضای خوشه‌های دیگر هیچ شباهتی ندارند.

الگوریتم K-Medoids جزء دسته الگوریتم‌های خوشه‌بندی مبتنی بر افراز داده‌ها است. در این نوع از تکنیک‌های خوشه‌بندی با فرض وجود n نمونه و ورودی k به عنوان تعداد خوشه‌های نهایی، نمونه‌ها در k خوشه گنجانده می‌شوند. معیاری جهت ارزیابی تشابه میان نمونه‌های یک خوشه و عدم تشابه میان خوشه‌ها برای شکل دهی آنها در حالت بهینه استفاده می‌شود. در این الگوریتم ابتدا باید k خوشه به صورت اتفاقی مشخص شود. پس از آن تعداد k نمونه به عنوان نماینده‌های^{۱۷} اولیه k خوشه (گاهی به آنها مرکز ثقل^{۱۸} یا مرکز خوشه نیز اطلاق می‌شود). به صورت اتفاقی انتخاب می‌شوند. پس از تشکیل ماتریس تشابه، هر یک از نمونه‌های باقیمانده (n-k نمونه) باید در یکی از این k خوشه قرار گیرند. به عبارت دیگر با کمک معیارهایی همچون فاصله‌ی اقلیدسی تشابه هر یک از نمونه‌های باقیمانده را با k نماینده محاسبه می‌کنیم و نمونه مورد نظر به هر یک نزدیکتر بود، به عضویت آن خوشه در می‌آید.

توجه کنید که می‌توانیم به جای تشکیل ماتریس تشابه، فاصله‌ی هر یک از نمونه‌های باقیمانده را با k نمونه اولیه محاسبه کنیم. هر نمونه به نزدیکترین نماینده تعلق دارد. پس از این با جایگزینی یک نمونه از داده‌ها با یکی از k نمونه نماینده، کیفیت و مناسب بودن خوشه‌های بدست آمده از این جایگزینی بررسی می‌شوند. در صورت بهبود در نتایج، مجاز به جایگزینی نماینده مزبور خواهیم بود. برای بررسی این موضوع از معیار خطای مطلق با رابطه زیر استفاده می‌شود:

$$E = \sum_{i=1}^k p_{ECi} |p - O_i| \quad (2-1)$$

در این رابطه E مجموع خطای مطلق برای کلیه نمونه‌ها، p یک نمونه از داده‌ها و O_i نماینده خوشه‌ی C_i است.

¹⁷ Centroid

¹⁸ Center Of Gravity

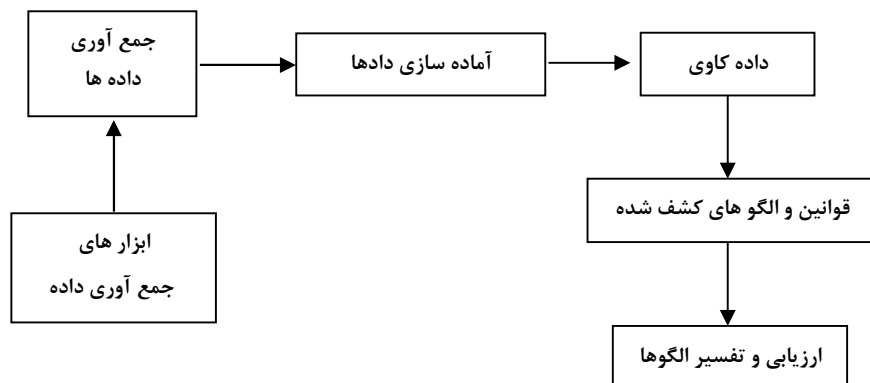
برای تعیین یک نماینده‌ی جدید O_{new} به جای نماینده کنونی O_{old} یک خوشه، عملیات زیر برای هر نمونه غیر نماینده مانند p باید انجام شود:

- اگر p متعلق به خوشه‌ای است که O_{old} نماینده آن خوشه باشد: چنانچه با جایگزینی O_{new} به جای O_{old} فاصله‌ی p به نماینده‌ی جدید یعنی O_{new} نزدیکتر باشد، p را در خوشه‌ای به نمایندگی O_{new} قرار دهید. در غیر اینصورت p را در خوشه‌ی i قرار دهید، به نحوی که فاصله‌ی نماینده آن خوشه با خودش حداقل باشد و i خواهد بود.

- اگر p متعلق به خوشه‌ی i است که O_i نماینده‌ی آن خوشه باشد: چنانچه با جایگزینی O_{new} به جای O_{old} همچنان فاصله‌ی p با نماینده‌ی خوشه خود یعنی O_i حداقل است، پس هیچگونه انتساب جدیدی صورت نمی‌پذیرد در غیر اینصورت p را در خوشه‌ای قرار می‌گیرد که نماینده آن O_{new} است دقت کنید که i خواهد بود. هر زمان که با جایگزینی O_{new} به O_{old} نمونه‌های p به خوشه‌ی دیگری منتقل می‌شوند، مقدار خطای مطلق نیز تغییر می‌کند. واضح است که کاهش در مقدار خطای مطلق، جایگزینی نماینده‌ی جدید O_{new} را تایید می‌کند. این فرآیند تا پوشش معیاری جهت خاتمه کار تکرار می‌شود. (تا هنگامی که دیگر هیچ یک از نمونه‌ها خوشه‌های خود را تغییر ندهند، ادامه پیدا می‌کند)

۲-۲- شرح مدل بکار رفته در مقاله

همانطور که ذکر شد هدف از این مقاله، ارائه روشی برای خوشه‌بندی داده‌های مصرف مشترکین آب با هدف شناسایی الگوی مصرف است. مدل اجرایی برای دستیابی به این هدف در شکل (۱) نمایش داده شده است.



شکل (۱) مدل اجرایی تحقیق

همانطور که در شکل (۱) مشاهده می‌کنید در این مدل ابتدا با جمع آوری اطلاعات مورد نیاز با کمک ابزارهایی مانند بانک اطلاعاتی رابطه ای و آماده سازی آن‌ها، زمینه داده کاوی فراهم می‌شود سپس، با کشف قوانین و الگوهای کشف شده در آن‌ها، اقدام به تفسیر و تحلیل نتایج با همکاری کارشناسان خبره می‌شود. گام‌های انجام این کار به شرح زیر می‌باشد:

گام اول: با کمک تکنیک پرس و جو، اطلاعات هر کاربری از انبار داده استخراج شد

گام دوم: انتقال داده‌های هر کاربری به نرم افزار

گام سوم: کد نویسی مربوط به الگوریتم K-Medoids

گام چهارم: مشخص نمودن تعداد خوشه‌ها

گام پنجم: اجرای الگوریتم

گام ششم: تهیه خروجی مناسب جهت مصور سازی و تهیه آمار

- ✓ مشخص نمودن اینکه هر مشترک در کدام خوشه قرار گرفته
 - ✓ مشخص نمودن مراکز خوشه‌ها در هر خوشه‌بند (الگوهای غالب در هر خوشه)
 - ✓ مشخص نمودن تعداد هر رکورد در هر شاخه و درصد گیری از آنها
- گام هفتم: تحلیل و ارزیابی بهترین خوشه‌بند برای هر کاربری

۳- ناحیه مورد مطالعه و مجموعه داده

از آنجا که عوامل متعددی از جمله فرهنگ و محدوده و شرایط آب و هوایی روی الگوی مصرف تاثیر به سزایی دارد لذا برای این تحقیق نیاز به انتخاب یک نمونه آماری بود به نحوی که جمعیت، کلیه شرایط تحقیق را داشته باشد و از طرفی زمینه ساز تعمیم و توسعه روش در کل شهرهای استان و ایران را نیز فراهم کند. با مطالعه پژوهش‌های قبلی و همچنین مذاکره با متخصصان حوزه مصرف مشترکین شرکت آب و فاضلاب استان فارس داده‌های مورد نیاز شناسایی شد. در این راستا از آنجایی که شهر فسا جزء ۴ شهر بزرگ استان فارس می‌باشد و دارای تنوع مشترکین در کاربری‌های مختلف است، پتانسیل قرار گرفتن به عنوان جامعه آماری را داشت. بعلاوه زیر ساخت‌های اطلاعاتی و همچنین دسترسی به منابع اطلاعاتی مورد نیاز در این شهر نیز به عنوان دیگر پارامترهای انتخاب این شهر به عنوان جامعه آماری بود. فسا شهری در ۱۴۵ کیلومتری جنوب شیراز واقع در استان فارس است. شهرستان فسا بین ۵۳ درجه و ۱۹ دقیقه تا ۵۴ درجه و ۱۵ دقیقه طول شرقی و ۲۸ درجه و ۳۱ دقیقه تا ۲۹ درجه و ۲۴ دقیقه عرض شمالی واقع شده است.

برای کشف رفتار مصرفی مشترکین آب شهر فسا می‌بایست کلیه اطلاعات (آخرین اطلاعات ثبت شده که به تایید نیز رسیده باشند) بررسی شوند. در نتیجه با استفاده از اسناد و مدارک موجود در حوزه مشترکین، کلیه اطلاعات مربوط به سال‌های ۹۰ تا ۹۲ شهر فسا که در سیستم‌های مشترکین ثبت شده بودند و از طرفی به تایید شرکت حسابرسی مربوطه نیز رسید بودند برای تحقیق انتخاب شد (اطلاعات مربوط به مشترکینی که دارای ۶ دوره قرائت کنتور در سال‌های ۹۰ تا ۹۲ می‌باشند در تمام کاربری‌ها استخراج شده‌اند). از آنجاکه این اطلاعات مورد استفاده مربوط به سال‌های گذشته شهر فسا می‌باشد و طبق اساسنامه شرکت در پایان هر سال می‌بایست کلیه اسناد و اطلاعات به تایید حسابررس قانونی شرکت برسد لذا اطلاعات بکار رفته از نظر روایی و پایایی در سطح بسیار بالایی قرار دارند.

با توجه به اینکه داده‌ها از بانک‌ها و منابع مختلف اطلاعاتی جمع آوری شده بود، در نتیجه کیفیت آن‌ها پایین بود، بنابراین در این مرحله برای دست یابی به مجموعه داده‌ای که بتوان از آن برای کشف الگو استفاده کرد (انبار داده)، ابتدا ویژگی‌های غیر مرتبط حذف و سپس با کمک توابع مختلف در محیط SQL عملیات تبدیل داده‌ها انجام شد. پس از پاک‌سازی، برای تعیین اینکه نتایج حاصله تا چه میزان دارای اعتبار هستند باید داده‌ها آزمون شوند. برای دست یابی به این هدف می‌بایست به بهترین شکل ممکن فاصله درون خوشه‌ای را کمینه و فاصله بین خوشه‌های را بیشینه نماید. به عبارت دیگر همه الگوریتم‌های خوشه‌بندی سعی در یافتن خوشه‌ها به گونه‌ای دارند که خوشه مشخص تر، مجزا تر و تفکیک شده تر باشند. دو فاکتور مهم برای این امر عبارتند از: تراکم خوشه‌ای و جدایی خوشه‌ای رابطه تراکم خوشه‌ای^{۱۹} SC و جدایی خوشه‌ای^{۲۰} S به شرح زیر می‌باشد

$$SC = \sum_{i=1}^c \frac{\sum_{k=1}^N (\mu_{i,k})^m \|x_k - v_i\|^2}{\sum_{k=1}^n (\mu_{i,k})^c - \sum_{j=1}^c \|v_j - v_i\|^2} \quad (3-1)$$

$$S = \frac{\sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^2 \|x_k - v_i\|^2}{N \min_{i,j} \|v_j - v_i\|^2} \quad (3-2)$$

¹⁹ Partition Index (SC)

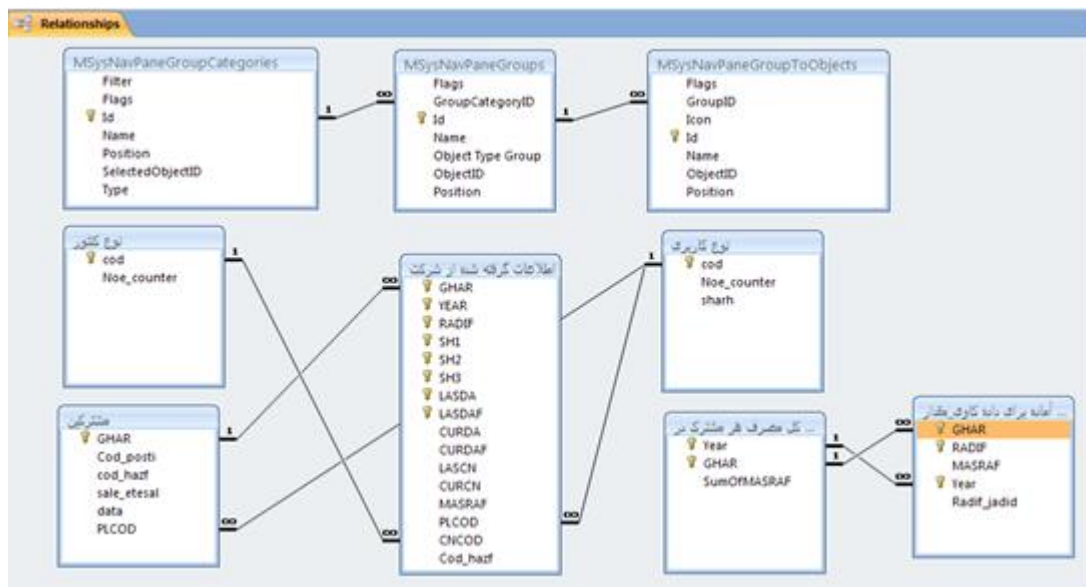
²⁰ Separation Index (S)

۴- مطالعه موردی:

همانطور که در بخش‌های قبل ذکر شد، هدف از این مقاله ارائه روشی برای خوشه‌بندی داده‌های مصرف مشترکین آب به منظور شناسایی الگوی مصرف آن‌ها است، بنابراین به دنبال کشف رابطه‌ها (الگوها) در جمع کثیری از داده‌ها می‌باشیم. با توجه به دارا بودن ویژگی‌های لازم، اطلاعات مربوط به مشترکین شهر فسا برای انجام مطالعه موردی انتخاب شد. تکنیک مورد استفاده برای دستیابی به این هدف نیز الگوریتم K-Medoids است، که در بخش ۲ به شرح کامل جزئیات این الگوریتم و نحوه پیاده‌سازی آن در این مقاله ذکر گردیده است. پیاده‌سازی این روش در دو مرحله انجام شده است بدین صورت که ابتدا داده‌های جمع‌آوری شده آماده‌سازی و سپس پیاده‌سازی الگوریتم انجام شده، در نهایت نیز نتایج حاصل از پیاده‌سازی این الگوریتم ارائه شده است. در ادامه به شرح کامل پیاده‌سازی این مساله پرداخته شده است.

۴-۱- آماده‌سازی داده‌ها

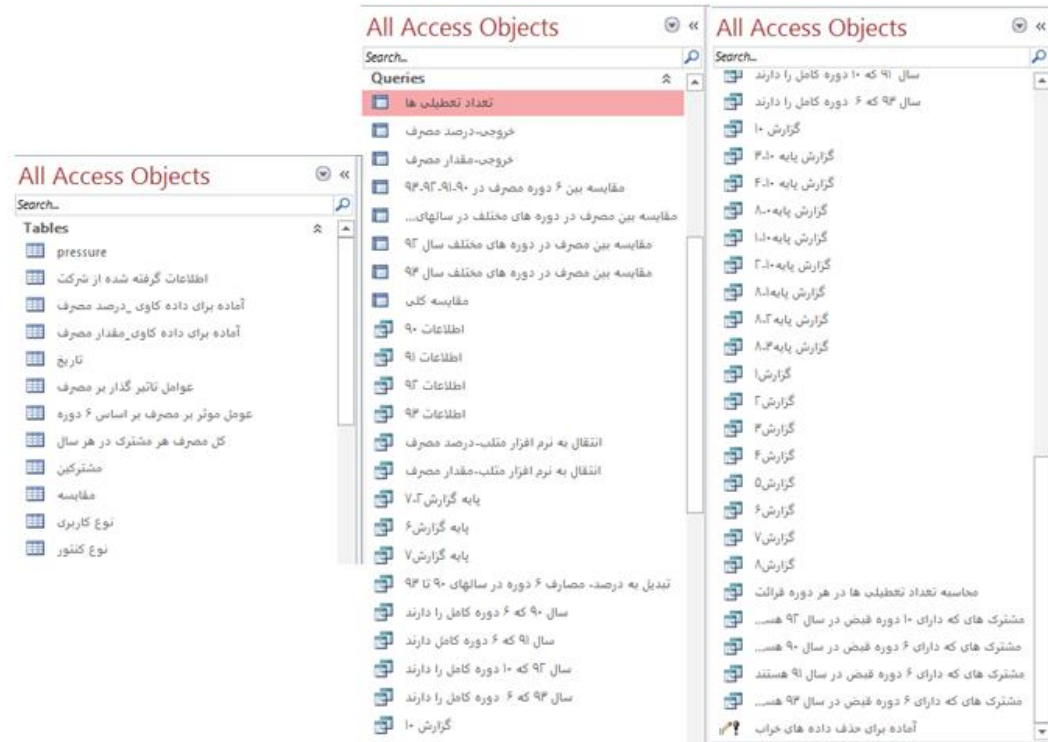
برای شناسایی الگوی مصرف مشترکین با همفکری با کارشناسان متخصص این حوزه فیلدهای اطلاعاتی مربوطه انتخاب شدند و متناسب با آن در محیط اکسس جداول مربوطه طراحی گردید. در انبار داده‌ها نام جدول "اطلاعات گرفته شده از شرکت" می‌باشد و این اصلی‌ترین جدول در بانک اطلاعاتی تهیه شده برای انبار داده تحقیق می‌باشد. رکوردهای این جدول بالغ بر ۱ میلیون می‌باشد که مربوط به مصرف تمام مشترکین شهر فسا در سال‌های ۹۰ تا ۹۲ است. آمار موجود قطعی و ثابت می‌باشد و به همین دلیل مبنای تحلیلی بالایی دارد. همانطور که گفته شد این جدول به عنوان هسته‌ی مرکزی ارتباطات در بانک اطلاعاتی می‌باشد و زمینه ایجاد بانک اطلاعاتی رابطه‌ای را فراهم می‌سازد، این بانک اطلاعاتی در شکل (۲) نشان داده شده است.



شکل (۲) بانک اطلاعاتی رابطه‌ای

در کنار این جدول، موجودیت دیگری تعریف می‌شود به نام کاربری که در آن انواع کاربری‌ها به همراه زیر مجموعه هر کدام مشخص می‌شود. کاربری‌های مشخص شده بر اساس نوع کنتور عبارتند از، مسکونی، تجاری، اداری، عمومی (به عنوان مثال بیمارستان‌ها، پارک‌ها و...)، حوزه دانشگاه، آموزش غیردولتی، امامزاده، گرمابه، نانوایی، نانوایی غیرسنتی، صنعتی (به عنوان مثال کارخانجات، کارگاه‌ها و ...)، فرهنگی دولت (به عنوان مثال کتابخانه‌ها، باشگاه‌ها و ...)، حوزه آموزش و پرورش، فرهنگی آزاد، خیریه، فضای سبز، غیرشرب، آزاد مسکونی، آزاد اداری، آزاد تجاری (منظور از آب آزاد، آبی است که جهت ساخت و ساز املاک دارای انشعاب دائم استفاده می‌شود)، خدمات

غیردولتی (به عنوان مثال روزنامه‌ها، دفاتر اسناد رسمی و ...)، روستا، نظامی، سکونت‌گاه دوم (به عنوان مثال ویلا، خانه باغ و ...)، امکان مذهبی و سایر. با استفاده از این اطلاعات و سایر ابزارها و تکنیک‌های موجود در نرم افزار اکسس اقدام به تهیه انبار داده، که به عنوان حیاتی‌ترین بخش داده کاوی است، شد. برای این منظور از کد برنامه نویسی به زبان SQL استفاده شده است. در انبار داده تهیه شده قسمت‌های زیادی را مشاهده می‌کنیم که همگی نقشی در آماده سازی داده‌ها برای داده کاوی داشته‌اند یا از آنها در هنگام پویس داده‌ها استفاده شده است. این انبار در شکل (۳) به نمایش درآمده است.



شکل (۳) اجزاء انبار داده

بعد از طراحی نقشه راه و اینکه چه اطلاعاتی مورد نیاز است اقدام به اجرای تحقیق شد. نتایج حاصل از اجرای تحقیق در بخش ۴-۲ ارائه شده است.

۴-۲- ارائه نتایج عددی

در این بخش، بر اساس مدل فرایندی مطرح شده در بخش‌های سوم، مراحل اجرایی روش تحقیق پیاده‌سازی شده و داده‌های تحقیق توصیف و تحلیل شدند و همچنین نحوه اعمال الگوریتم داده کاوی و تجزیه و تحلیل انجام شده روی مجموعه داده‌های واقعی شرح داده شده است. سپس نتایج حاصل از ساختار اجرایی و پیاده سازی روش تحقیق آورده شده و در نهایت نتایج و خروجی‌های حاصله از آن ارائه شده است. در ادامه به شرح کامل این موضوع پرداخته شده است.

پس از اجرای الگوریتم توسط نرم افزار متلب و محاسبه معیارهای ارزیابی مشخص شد که هر چه تعداد خوشه‌ها زیادتر باشد پراکندگی و فشردگی نیز بیشتر خواهد شد (دو شاخص ارزیابی کمتر خواهند شد)، اما در مباحث تئوری خوشه‌بندی همیشه هدف کم کردن خوشه‌ها است تا بهتر بتوان مدیریت برنامه ریزی و کنترل را انجام داد. نتایج مربوط به خوشه‌بندی‌های انواع کاربری‌ها بر اساس معیارهای SC و S در جدول (۱) ارائه شده است، همچنین این نتایج در شکل (۴) نمایش داده شده است.

جدول (۱) ارزیابی معیارهای SC و S برای انواع کاربری‌ها

S	SC	خوشه‌بندی	نوع کاربری	S	SC	خوشه‌بندی	نوع کاربری
0.00000568	0.0088634	خوشه ۲	آزاد مسکونی	0.00000015	0.0080091	خوشه ۲	مسکونی
0.00000159	0.0020034	خوشه ۳		0.00000024	0.0080479	خوشه ۳	
0.00000262	0.0024586	خوشه ۴		0.00000005	0.0015144	خوشه ۴	
0.00000210	0.0019989	خوشه ۵		0.00000007	0.0021971	خوشه ۵	
0.00000233	0.0020035	خوشه ۶		0.00000007	0.0024205	خوشه ۶	
0.00000197	0.0018087	خوشه ۷		0.00000005	0.0014598	خوشه ۷	
0.00000193	0.0014049	خوشه ۸*		0.00000005	0.0012450	خوشه ۸*	
0.00000265	0.0022689	خوشه ۹		0.00000005	0.0012654	خوشه ۹	
0.00000209	0.0016918	خوشه ۱۰		0.00000005	0.0012944	خوشه ۱۰	
0.00000211	0.0017358	خوشه ۱۱		0.00000004	0.0011068	خوشه ۱۱	
0.00000228	0.0018310	خوشه ۱۲		0.00000005	0.0012620	خوشه ۱۲	
0.00000146	0.0011218	خوشه ۱۳		0.00000004	0.0011462	خوشه ۱۳	
0.00000162	0.0013000	خوشه ۱۴		0.00000004	0.0010753	خوشه ۱۴	

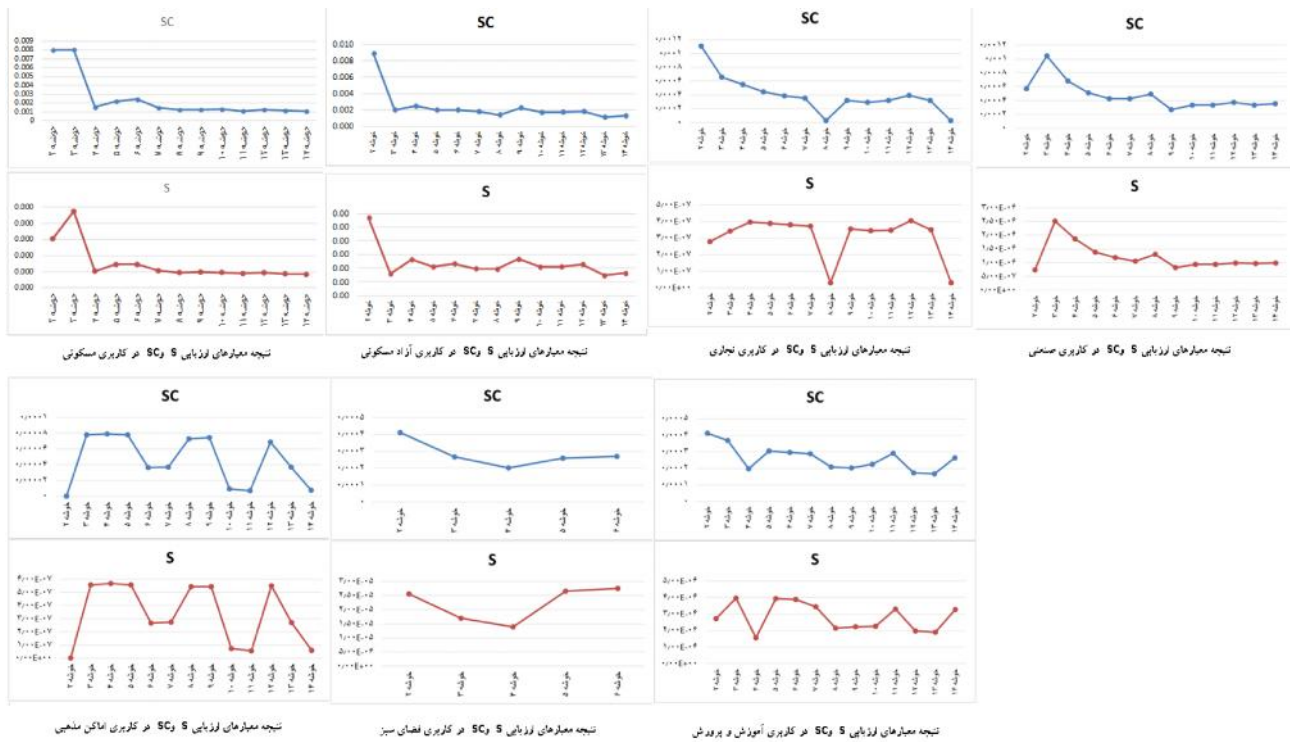
S	SC	خوشه‌بندی	نوع کاربری	S	SC	خوشه‌بندی	نوع کاربری
0.00000075	0.0005753	خوشه ۲	صنعتی	0.00000028	0.0011078	خوشه ۲	تجاری
0.00000251	0.0010459	خوشه ۳		0.00000034	0.0006604	خوشه ۳	
0.00000187	0.0006857	خوشه ۴		0.00000040	0.0005500	خوشه ۴	
0.00000138	0.0005130	خوشه ۵		0.00000039	0.0004458	خوشه ۵	
0.00000118	0.0004314	خوشه ۶		0.00000038	0.0003885	خوشه ۶	
0.00000105	0.0004304	خوشه ۷		0.00000037	0.0003530	خوشه ۷	
0.00000130	0.0004970	خوشه ۸		0.00000003	0.0000310	خوشه ۸*	
0.00000083	0.0002711	خوشه ۹*		0.00000036	0.0003176	خوشه ۹	
0.00000094	0.0003389	خوشه ۱۰		0.00000034	0.0002965	خوشه ۱۰	
0.00000093	0.0003376	خوشه ۱۱		0.00000035	0.0003208	خوشه ۱۱	
0.00000099	0.0003753	خوشه ۱۲		0.00000041	0.0003921	خوشه ۱۲	
0.00000098	0.0003377	خوشه ۱۳		0.00000035	0.0003182	خوشه ۱۳	
0.00000098	0.0003563	خوشه ۱۴		0.00000003	0.0000278	خوشه ۱۴	

ادامه جدول (۱) ارزیابی معیارهای SC و S برای انواع کاربری‌ها

S	SC	خوشه‌بندی	نوع کاربری	S	SC	خوشه‌بندی	نوع کاربری
0.00000273	0.0004115	خوشه ۲	آموزش و پرورش	0.00000000	0.00000000	خوشه ۲	اماکن مذهبی
0.00000397	0.0003673	خوشه ۳		0.00000056	0.0000778	خوشه ۳	
0.00000157	0.0001961	خوشه ۴*		0.00000057	0.0000791	خوشه ۴	
0.00000395	0.0003055	خوشه ۵		0.00000056	0.0000779	خوشه ۵	
0.00000388	0.0002978	خوشه ۶		0.00000027	0.0000364	خوشه ۶*	
0.00000344	0.0002870	خوشه ۷		0.00000028	0.0000372	خوشه ۷	
0.00000214	0.0002086	خوشه ۸		0.00000055	0.0000725	خوشه ۸	
0.00000222	0.0002020	خوشه ۹		0.00000054	0.0000746	خوشه ۹	
0.00000226	0.0002242	خوشه ۱۰		0.00000007	0.0000094	خوشه ۱۰	
0.00000331	0.0002918	خوشه ۱۱		0.00000006	0.0000068	خوشه ۱۱	
0.00000197	0.0001722	خوشه ۱۲		0.00000055	0.0000689	خوشه ۱۲	
0.00000191	0.0001666	خوشه ۱۳		0.00000027	0.0000367	خوشه ۱۳	
0.00000328	0.0002639	خوشه ۱۴		0.00000006	0.0000073	خوشه ۱۴	

S	SC	خوشه‌بندی	نوع کاربری
0.00002566	0.0004106	خوشه ۲	فضای سبز
0.00001707	0.0002664	خوشه ۳	
0.00001397	0.0002016	خوشه ۴*	
0.00002668	0.0002600	خوشه ۵	
0.00002761	0.0002716	خوشه ۶	

در جدول (۱) نتایج مربوط به خوشه‌بندی برای انواع کاربری‌ها آورده شده و تعداد بهینه خوشه‌بند برای هر کاربری بر اساس معیارهای SC و S متمایز شده است. بر طبق جدول (۱) برای کاربری مسکونی از بین خوشه‌بندها، خوشه‌بندهای ۴، ۵، ۷، ۸ کاندید هستند که در بین خوشه‌بندهای کاندید، خوشه‌بند ۸ که مقادیر SC و S کمتری دارد، انتخاب می‌شود. برای کاربری روستایی و حوزه دانشگاه، با توجه به تعداد کم این نوع کاربری، بیش از ۲ خوشه‌بند در در الگوریتم جوابی وجود نداشت، بنابراین همان خوشه‌بند ۲ انتخاب شد. برای کاربری تجاری، در خوشه‌بندهای ۸ و ۱۴ کمترین مقدار برای شاخص‌های ارزیابی مشاهده شد و از آنجایی که هر چه قدر تعداد خوشه‌بندها کمتر باشد وضعیت بهتر است، خوشه‌بندی ۸ انتخاب می‌شود. برای اماکن مذهبی نیز پس از اجرای الگوریتم توسط نرم افزار متلب و محاسبه معیارهای ارزیابی مشخص شد که کمترین مقدار شاخص‌های ارزیابی در خوشه‌بند ۱۰ و ۱۱ است، اما از آنجایی که به دنبال کم کردن تعداد خوشه‌بندها هستیم به سراغ دیگر خوشه‌بندها می‌رویم. خوشه‌بند ۶ و ۷ کاندیدهای بعدی برای این نوع کاربری هستند که از بین آن‌ها خوشه‌بند ۶ که معیارهای کمتری دارد انتخاب می‌شود. برای کاربری‌های آزاد مسکونی، صنعتی، فضای سبز و آموزش نیز به ترتیب ۸، ۹، ۴ و ۴ خوشه‌بند انتخاب شد.



شکل (۴) نتیجه معیارهای ارزیابی SC و S

پس از تعیین خوشه‌بندها برای تمامی انواع کاربری‌ها، نتایج مربوط به الگوی مصرف مشترکین انواع کاربری‌ها در جدول (۲) ارائه شده است.

جدول (۲) خروجی نرم‌افزار متلب برای الگوی مصرف انواع کاربری‌ها

درصد اعضای خوشه (در هر کاربری)	دوره‌ها						مراکز خوشه (در هر کاربری)	نوع کاربری
	بهمن و اسفند	آذر و دی	مهر و آبان	مرداد و شهریور	خرداد و تیر	فروردین و اردیبهشت		
10.2%	49	53	71	80	68	58	خوشه ۱	مسکونی
0.9%	21	26	29	49	37	187	خوشه ۲	
11.9%	8	9	10	12	9	11	خوشه ۳	
24.0%	16	20	23	27	23	22	خوشه ۴	
27.9%	24	27	35	41	35	32	خوشه ۵	
0.2%	21	22	39	320	5	10	خوشه ۶	
24.9%	33	41	49	55	45	42	خوشه ۷	
0.1%	17	308	10	15	5	11	خوشه ۸	
40%	54000	52000	52000	20770	13087	21200	خوشه ۱	روستایی
60%	8949	6950	7000	6132	8274	12000	خوشه ۲	

ادامه جدول (۲) خروجی نرم افزار متلب برای الگوی مصرف انواع کاربری ها

16%	25	30	32	34	29	27	خوشه ۱	آزاد مسکونی
17%	38	48	67	66	63	46	خوشه ۲	
1%	195	5	15	20	10	5	خوشه ۳	
42%	5	5	5	5	4	6	خوشه ۴	
4%	0	18	23	21	26	61	خوشه ۵	
8%	5	4	16	5	5	5	خوشه ۶	
1%	105	107	120	200	221	245	خوشه ۷	
11%	5	10	10	25	12	7	خوشه ۸	
11%	104	129	231	323	421	390	خوشه ۱	حوزه دانشگاه
89%	40	40	100	60	75	55	خوشه ۲	
23.4%	5	6	11	16	14	10	خوشه ۱	کاربری تجاری
0.3%	485	330	455	462	366	336	خوشه ۲	
10.5%	22	27	41	46	33	28	خوشه ۳	
0.1%	0	72	0	0	4412	5929	خوشه ۴	
0.9%	9	19	6	11	23	122	خوشه ۵	
2.2%	35	25	20	15	3	2	خوشه ۶	
2.2%	76	73	108	121	130	76	خوشه ۷	
60.4%	1	1	2	3	3	2	خوشه ۸	
6%	13	16	39	20	14	11	خوشه ۱	صنعتی
9%	9	6	16	14	13	20	خوشه ۲	
22%	25	29	35	41	40	39	خوشه ۳	
7%	160	160	226	323	244	217	خوشه ۴	
1%	321	488	427	681	894	719	خوشه ۵	
9%	13	20	13	35	16	13	خوشه ۶	
15%	2	2	3	2	2	2	خوشه ۷	
21%	54	59	99	100	85	60	خوشه ۸	
11%	5	9	10	10	7	3	خوشه ۹	
46%	20	40	40	50	40	40	خوشه ۱	اماکن مذهبی
1%	1	2	9953	15	0	0	خوشه ۲	
9%	167	201	187	188	252	127	خوشه ۳	
1%	27	31	7	9534	0	603	خوشه ۴	
3%	60	481	691	778	533	697	خوشه ۵	
40%	5	15	10	15	10	20	خوشه ۶	
13%	170	1615	1471	816	558	110	خوشه ۱	فضای سبز
13%	546	1920	227	265	323	125	خوشه ۲	
13%	1790	906	632	1020	987	169	خوشه ۳	
63%	378	185	360	366	180	232	خوشه ۴	
68%	65	72	94	70	69	75	خوشه ۱	حوزه آموزش و پرورش
2%	425	941	1191	1331	1653	1262	خوشه ۲	
27%	142	204	257	261	243	177	خوشه ۳	
3%	281	439	765	599	681	1496	خوشه ۴	

همانطور که در جدول (۲) مشاهده می‌کنید، درصد توزیع مشترکین برای هر کاربری مشخص شده است، که بر این اساس شرکت می‌کند برنامه‌ریزی‌هایی در راستای هر کدام از کاربری‌ها داشته باشد. تحلیل‌های صورت گرفته بر روی نتایج به شرح زیر می‌باشد:

- کاربری مسکونی:
در کاربری مسکونی، ۲۴٪ مشترکین کاربری در خوشه ۷ قرار گرفته‌اند، بنابراین شرکت می‌تواند با بکارگیری تکنیک‌های مدیریت مصرف (مانند نصب تجهیزات کاهنده) مشترکین این خوشه را به خوشه‌ایی با مصرف پایین‌تر منتقل نماید.
- کاربری روستایی:
در کاربری روستایی، ۴۰٪ مشترکین در خوشه ۱ قرار دارند و مصرف بالایی دارند، که این میزان مصرف نشان می‌دهد شرکت می‌بایست ضمن بررسی دقیق جمعیت این مشترکین نسبت به ارائه راهکار برای کاهش مصرف آن‌ها اقدام نماید.
- کاربری آزاد مسکونی:
با دقت در خروجی حاصل شده کاربری آزاد مسکونی مشاهده می‌شود که دامنه مصرف (تقریباً) در تمام دوره‌ها ۵ مترمکعب است، بنابراین جا دارد که حوزه مشترکین آب و فاضلاب روی این مساله بررسی بیشتری را انجام دهند.
- کاربری حوزه دانشگاه:
بر اساس خروجی مربوط به حوزه دانشگاه نیز مشخص می‌شود که تقریباً ۹۰٪ مشترکین در خوشه ۲ قرار دارند، که مصرف در این خوشه به خصوص در ۲ دوره آخر سال یکسان است، بنابراین به نظر می‌رسد حوزه مشترکین آب و فاضلاب می‌بایست مصرف مشترکینی که در این خوشه قرار می‌گیرند را زیر نظر گرفته و تحلیل نمایند.
- کاربری تجاری:
نتایج مربوط به این نوع کاربری نشان می‌دهد که ۶۰٪ مشترکین در خوشه ۸ قرار دارند و این در حالی است که دامنه مصرف به خصوص در ۲ دوره آخر سال یکسان است. پیشنهاد می‌شود حوزه مشترکین با کمک تکنیک‌های آب بدون درآمد در بخش هدر رفت ظاهری نسبت به نظارت و کنترل مصرف این گروه از مشترکین اقدام نمایند.
- کاربری صنعتی:
۱۵٪ مشترکین این کاربری مصرفی شبیه خوشه ۷ دارند که روند مصرف این خوشه نشان‌دهنده یک مشکل در این گروه می‌باشد و جا دارد شرکت نسبت به بررسی و نظارت اینگونه مشترکین هرچه سریعتر اقدام نماید.
- کاربری مذهبی:
با دقت در مصرف خوشه‌های این کاربری مشخص می‌شود که ۴۶٪ مشترکینی که در خوشه ۱ قرار گرفته‌اند مصرف منطقی ندارند و می‌بایست مصرف آن‌ها کنترل شود. همچنین در سایر خوشه‌ها نیز الگوی مصرف مشخص شده خیلی منطقی به نظر نمی‌رسند.
- کاربری فضای سبز:
خروجی جدول مربوط به کاربری فضای سبز نشان می‌دهد که مشکل مصرف بخصوص در خوشه‌های ۳ و ۱ است، بنابراین حوزه مشترکین شرکت می‌بایست دقت نظر بیشتری روی مصرف این مشترکین داشته باشند.
- کاربری آموزش و پرورش:
با دقت در دانش استخراجی از داده‌ها برای این کاربری مشاهده می‌شود که ۶۸٪ مشترکین در این نوع کاربری در خوشه ۱ قرار دارند، که دقت در عملکرد مصرف آن‌ها نشان‌دهنده این موضوع است که حوزه مشترکین شرکت می‌بایست دقت نظر بیشتری بر این گونه مشترکین داشته باشد.

۵- نتیجه‌گیری و ارائه پیشنهادات تحقیقات آتی

پرسش اصلی مطرح شده این بود که بوسیله یافته‌های داده کاوی، شرکت آب و فاضلاب چه الگوها و مشکلاتی را می‌تواند در جهت کنترل منابع آب شهری و بهینه‌سازی توزیع آب، شناسایی کند تا بتواند در آینده به وسیله اقدامات و تدابیر و تصمیمات لازم مدیریتی به رفع آنها پردازد؟ برای پاسخ به این سئوالاتی از اینگونه، از روش خوشه‌بندی و الگوریتم K-Medoids استفاده شد و به تفکیک

الگوهای مصرف هر کاربری به تفکیک مورد بررسی و تحلیل گردید. نتایج حاصل از اجرای الگوریتم نشان داد که، الگوهای مصرف در اکثر کاربری‌ها نیاز به بررسی و نظارت بیشتر دارد همچنین با کمک این دانش کسب شده، زمینه پیاده سازی تکنیک های مدیریت مصرف در کاربری های مختلف و خوشه های گوناگون قابل برنامه ریزی بوده و مسیر جهت دستیابی به اهداف این تکنیک ها میسرتر خواهد بود. به عنوان زمینه‌هایی برای تحقیقات آتی پیشنهاد می‌گردد، با همکاری شرکت‌های خدماتی دیگر مثل برق و گاز می‌توان نسبت به بررسی الگوی مصرف مشترکین در خدمات مختلف اقدام نمود. شاید این موضوع یکی از جدیدترین مواردی که می‌توان نسبت به شناسایی الگوی مصرف هر عضو (مشترک) در زمینه‌های خدماتی مثل آب، برق و گاز در جهان انجام داد. همچنین چنانچه سیستم GIS موجود باشد، می‌توان میزان خوشه‌بندی و شناسایی الگوها را منطبق بر مناطق جغرافیایی انجام داد و در نهایت چگالی مصرف و همچنین الگوی مصرف منطقه‌ای را استخراج نمود. بعلاوه می‌توان از سایر روش‌ها و الگوریتم‌های داده‌کاوی از جمله الگوریتم درخت تصمیم (به منظور پیش بینی مقدار مصرف مشترکین و بررسی وضعیت کنتور خراب آنها)، روش رگرسیون (به منظور در نظر گرفتن متغیرهای تاثیرگذار در مصرف مشترکین و ارتباط بین آنها)، روش تشخیص انحراف (شناسایی مشترکین دارای انشعابات غیرمجاز، یا دارای کنتورهای خراب و خطای زیاد) و روش خوشه‌بندی مبتنی بر چگالی (به منظور شناسایی میزان مصرف در مناطق مختلف در جهت برنامه‌ریزی به توزیع آب در سطح شهر) استفاده نمود.

۵- مراجع

- [1] مبینی، ع.، ۱۳۸۶، راهنمای مدیریت تقاضا (مصرف) در سیستم آب و فاضلاب، شرکت مهندسی آب و فاضلاب کشور
- [2] Berry, M., and Linoff, G., 2012, "Data Mining Techniques: For Marketing, Sales, and Customer Support". 3rd edition. New York: John Wiley and Sons
- [3] *An Introduction to Data Mining*: <http://www.theartling.com/>, retrieved on December 2, 2014.
- [4] طباطبایی، ا.، نگرشی بر داده کاوی، دانشگاه آزاد قزوین، ۱۳۸۸، ص ۱۰-۱۵
- [5] Babovic, V., Drécourt, J. P., Keijzer, M., & Hansen, P. F. (2002). A data mining approach to modelling of water supply assets. *Urban Water*, 4(4), 401-414.
- [6] Giustolisi, O., Savic, A., Laucelli, D., 2004, *Data Mining for Management and Rehabilitation of Water Systems: The Evolutionary Polynomial Regression Approach*, Wasserbauliche Mitteilungen.
- [7] Díaz, J. L., Herrera, M., Izquierdo, J., & Pérez-García, R. (2010). *The tasks of pre and post-processing in data mining applied to a real world problem* (Doctoral dissertation, International Environmental Modelling and Software Society).
- [8] Boyle, C. E., Eskaf, S., Tiger, M. W., & Hughes, J. A. (2011). Mining water billing data to inform policy and communication strategies. *Journal AWWA*, 103(11), 45-58.
- [9] R Cardell-Oliver, G Peach, 2013, *a data mining approach for discovering Water use Pattern*, Smart metering Data, 2-6
- [10] Dutta, P., 2012, *A survey Of Data Mining Applications in Water Quality Management*, CUBE International Information Technology Conference, USA, 470-475
- [11] Khan, M. A., Islam, M. Z., & Hafeez, M. (2012, December). Evaluating the performance of several data mining methods for predicting irrigation water requirement. In *Proceedings of the Tenth Australasian Data Mining Conference-Volume 134* (pp. 199-207). Australian Computer Society, Inc...
- [12] Wen, Y., Huang, W., Wu, J., 2013, *Water Consumption analysis System Based On data Mining*, Mechanics and Materials, 241-244, 1093-1097
- [13] Sun, J., Wang, R., Wang, X., Ping, J., 2014, *Spatial cluster analysis of bursting pipes in water supply networks*, Procedia Engineering ,70,1610 – 1618
- [14] آقابابایی، ع.، شهرابی، ج.، کاربرد دانش داده کاوی در مواجهه با مشترکین بدهکار در شرکت آب و فاضلاب مشهد (منطقه ۳) ، ششمین کنفرانس داده کاوی، ۱۳۹۱
- [15] Miyamoto, S., Ichihashi, H., & Honda, K. (2008). Algorithms for fuzzy clustering. *Methods in c-Means Clustering with Applications*. Kacprzyk J, editor Berlin: Springer-Verlag.