طراحی مدل داده‌کاوی تلفیقی خوشه‌بندی-وابستگی جهت بررسی رفتار مصرف برق واحدهای صنعتی

فرزاد رحیمی موگویی
دانشجوی دکتری مدیریت سیستم‌ها، دانشگاه سمنان
Rahimi.Farzad@semnan.ac.ir

رضا کامران راد (نویسنده مسئول)
استادیار مهندسی صنایع، دانشگاه سمنان
R.Kamranrad@semnan.ac.ir

عظیم‌اله زارعی
استاد مدیریت، دانشگاه سمنان
A_Zarei@semnan.ac.ir

تاریخ دریافت:
۱۴۰۱/۳/۸

تاریخ پذیرش:
۱۴۰۱/۱۲/۱۶

کلمات کلیدی:
داده‌کاوی
خوشه‌بندی
تجزیه و تحلیل سلسله مراتبی
الگوریتم کا-میانگین
قوانین وابستگی
بهینه‌سازی مصرف انرژی

چکیده

با توجه به سهم بالای مصرف برق در صنایع کشور، طی چند سال اخیر طرح‌های مختلفی از جمله خاموشی‌های سراسری در اوقات پیک مصرف اجرا شده است. امروزه داده‌کاوی به عنوان فرآیند کشف الگوهای مفید از پایگاه داده و یکی از روش‌های مؤثر برای تجزیه و تحلیل، مدل سازی و پیش‌بینی مصرف انرژی کاربرد فراوانی پیدا کرده است. در این مطالعه، مدلی تلفیقی جهت بررسی رفتار مصرف برق با استفاده از تکنیک‌های خوشه‌بندی کا-میانگین و قوانین وابستگی جهت کشف و استخراج الگو از مجموعه داده‌های مربوط به مصرف برق واحدهای صنعتی مستقر در یکی از شهرک‌های صنعتی استان تهران طراحی شده است. مشاهدات نشان می‌دهد که طی ماه‌های گرم سال، میانگین مصرف واحدهای خوشه پر-مصرف که حدود ۳۴ درصد واحدهای صنعتی مورد مطالعه را شامل می‌شود، حدود ۴,۲ برابر مصرف خوشه کم‌مصرف و حدود ۱,۷ برابر مصرف خوشه متوسط است. با بکارگیری مدل پیشنهادی در این پژوهش ضمن شناسایی واحدهای پرمصرف و اعمال سیاست‌های هوشمندانه و عادلانه در خاموشی اجباری، می‌توان علاوه بر تشویق واحدهای صنعتی به بهینه‌سازی مصرف انرژی، از ایجاد خسارت ناشی از توقف‌های اجباری تولید ممانعت کرد. رویکرد نوآورانه این مدل قادر به کنترل حجم زیادی از داده‌ها برای برنامه‌ریزی مناطق مختلف با هدف بهینه‌سازی در مصرف انرژی آن می‌باشد.

۱ مقدمه

۱-۱- وضعیت مصرف انرژی الکتریکی بخش صنعت

با توجه به پیشرفت جوامع صنعتی و استفاده روزافزون از تجهیزات و ماشین‌آلات با مصرف انرژی الکتریکی در کلیه بخش‌های مختلف صنعتی، تجاری، کشاورزی و خانگی، مصرف برق امری اجتناب ناپذیر است. افزایش هزینه‌های انرژی و سهم رو به افزایش آن در هزینه تمام شده تولید محصولات صنعتی، چالش مهمی برای صاحبان صنایع محسوب می شود. از طرفی محدودیت‌های ظرفیت تولید برخی از انرژی‌های استراتژیک (مانند برق) و آلایندگی ناشی از تولید انرژی از سوخت‌های فسیلی، مصرف انرژی را به یک مسئله جهانی تبدیل نموده که دولت‌ها برای کاهش و مصـــرف بهینه آن در اشـــکال مختلف اقداماتی را برنامه‌ریزی و اجرا می‌نمایند. با بررســـی ترازنامه انرژی کشور در سال ۱۳۹۷ (جدول ۱) مشخص می‌شود سهم بخش صنعت از کل مصرف برق کشور حدود ۳۴ درصد (معادل ۸۸,۵ تراوات ساعت) بوده که بیشترین سهم از مصرف برق کشور را در بین کلیه بخش‌ها به خود اختصاص داده و نسبت به سال قبل ۵,۱ درصد افزایش داشته است. همچنین سرانه مصرف نهایی انرژی کشور در بخش صنعت ۱,۵ برابر متوسط جهانی است که این امر از بهره‌وری (نسبت بین خروجی به ورودی) پایین در بهره‌برداری، مصرف بالای انرژی و همچنین استفاده از کالاها و خدمات انرژی‌بر ناشی می شود. بررسی سایر شاخص‌های مرتبط با مصرف انرژی نیز بر بهره‌وری پایین انرژی در کشور تأکید می‌کند. از جمله این شاخص‌ها، شاخص شدت انرژی است که برای تعیین کارایی انرژی در سطح اقتصاد ملی هر کشور استفاده می شود که از تقسیم مصرف نهایی انرژی بر تولید ناخالص داخلی محاسبه می‌شود و نشان می‌دهد که برای تولید مقدار معینی از کالاها و خدمات، چه مقدار انرژی به کار رفته است. داده‌های آماری مصرف انرژی نشان می‌دهد شدت مصرف انرژی در کشور در سال ۲۰۱۷ بیش از ۲,۶ برابر متوسط جهانی است (ترازنامه انرژی سال ۱۳۹۷، ۱۳۹۹).

تعداد جمعیت، حجم فعالیت‌های صنعتی و اقتصادی و وضعیت آب و هوا از عوامل تأثیرگذار بر مصرف برق استان‌های کشور است، به گونه ای که استان تهران با مصرف ۳۴۸۴۷,۲ گیگاوات ساعت برق به تنهایی ۱۳,۴ درصد از برق مصرفی تأمین شده در کشور را به مصرف می رساند. در بخش صنعت نیز استان اصفهان با ۱۴۴۰۸,۹ گیگاوات ساعت بیشترین میزان مصرف برق را به خود اختصاص داده است. در جدول ۲، درصد رشد سالانه مصرف انرژی در بخش صنعت به تفکیک حامل‌های انرژی طی سال‌های ۱۳۸۹ تا ۱۳۹۷ آمده و بیانگر رشد ۳۵ درصدی در کل مصرف حامل‌های انرژی و ۴۳ درصدی در مصرف برق طی این مدت می‌باشد که به صورت میانگین رشد سالانه حدود ۴,۵ درصدی در کل مصرف حامل‌های انرژی و حدود ۵,۴ درصدی در مصرف برق را نشان می‌دهد (ترازنامه انرژی سال ۱۳۹۷، ۱۳۹۹).

جدول (۱): درصد سهم بخش‌های مختلف از مصرف برق کشور در سال ۱۳۹۷

بخش	صنعت	خانگی	تجاری و عمومی	کشاورزی	سایر مصارف	حمل و نقل
درصد سهم	۳۴	۳۲,۸	۱۶,۶	۱۴,۵	۱,۹	۰,۲

جدول(۲): درصد رشد سالانه مصرف انرژی بخش صنعت به تفکیک حامل‌های انرژی (ترازنامه انرژی سال ۱۳۹۷، ۱۳۹۹)

حامل انرژی	۱۳۸۹	۱۳۹۰	۱۳۹۱	۱۳۹۲	۱۳۹۳	۱۳۹۴	۱۳۹۵	۱۳۹۶	۱۳۹۷
فرآورده‌های نفتی	-۱۰,۴۸	-۳۴,۶۵	۶,۹۸	-۱۱,۳۷	۰,۶۳	-۲۸,۵۳	-۳,۲۶	۱,۲۹	۳,۱۴
گاز طبیعی	۱۷,۷۶	۱۱,۶۳	۶,۱۵	۲,۹۹	۵,۹۷	-۲,۶۸	۷,۳۹	۳,۹۵	۶,۷۵
زغال سنگ	-۷۱,۶۲	۷,۸۶	-۲,۴۶	۰	۴۳,۳۲	-۱۸,۳۳	۲۱,۲۷	-۳,۲۹	-۱۳,۱۴
برق	۸,۰۴	۹,۶۵	۴,۱۲	۰,۰۷	۵,۵۴	۱,۷۶	۴,۵۶	۷,۱۰	۲,۰۷
کل مصرف انرژی	**۹,۳۸**	**۱,۹۸**	**۵,۹۷**	**۰,۹۱**	**۵,۳۸**	**-۴,۹۶**	**۶,۱۳**	**۴,۱۹**	**۵,۷۲**

۱-۲- داده‌کاوی

داده کاوی فرآیند استفاده از یک یا چند تکنیک مبتنی بر رایانه برای استخراج اطلاعات ضمنی، ناشناخته، بالقوه مفید و استخراج دانش از داده‌ها است (کالایسوی[۱] و عبدالسمات[۲]، ۲۰۱۹). امروزه داده‌کاوی به عنوان یکی از ابزارهای بسیار مهم مدیران جهت شناخت و ضعیت دقیق سازمان و همچنین کمک در اتخاذ تصمیات مناسب، کاربرد دارد. با استفاده از این روش، داده‌های موجود در سازمان با بکارگیری ابزارهای نرم‌افزاری مورد بررسی و تحلیل دقیق قرار می‌گیرد تا الگوهای پنهان و پیچیده‌ای که در آن‌ها وجود دارد، کشف و استخراج شود (آذر و خدیور، ۱۳۹۸). هدف اصلی داده‌کاوی می‌تواند توصیف یا پیش‌بینی است و می‌توان گفت داده‌کاوی شناسایی الگوهای صحیح، بدیع، سودمند و قابل درک از داده‌های موجود در یک پایگاه داده است که با استفاده از پردازش‌های معمول قابل دستیابی نیستند (قره‌خانی و ابوالقاسمی، ۱۳۹۰). به عبارت دیگر، داده‌کاوی فرآیند به خدمت گرفتن یک روش‌شناسی رایانه‌ای است که با استفاده از روش‌ها و الگوریتم‌های مختلف، در جستجوی دانش نهفته در داده‌هاست (کانتارزیچ[۳]، ۲۰۱۹). موضوعی که در بسیاری از پژوهش‌ها کمتر به آن توجه شده است، تمرکز بر جنبه‌های فنی روش‌های داده‌کاوی و غفلت از نتایج کاربردی و مطابق با واقعیت‌های محیط مورد بررسی است. این امر ناشی از آن است که به دلیل پیچیدگی‌های نسبی روش‌های داده‌کاوی، بیشتر وقت و انرژی پژوهشگران بر حل مسائل تکنیکی متمرکز می‌شود. به علاوه کاربرد روش‌های گوناگون داده‌کاوی ممکن است به نتایج متفاوتی نیز بیانجامد و لازم است پژوهشگران نتایج روش‌های مختلف داده‌کاوی را مقایسه و آن‌ها را در کنار شرایط محیط مورد مطالعه، به کار گیرند (لیتراس[۴] و همکاران، ۲۰۲۰).

۲- مرور ادبیات و تحقیقات پیشین

بر اساس تجزیه و تحلیل‌های تحت نظارت انجام شده در پژوهشی (کریم‌تبار و همکاران، ۲۰۱۵)، مدل رگرسیون نرخ خطای[۵] کمتری را در مقایسه با سایر مدل‌ها در رابطه با پیش‌بینی مصرف برق تولید می‌کند. متغیرهای پیش‌بینی کننده در این تحقیق شامل ساکنان، دما، رطوبت و قیمت مصرف برق است. بر اساس این مطالعه، میزان رشد مصرف برق پس از گذشت ۷ سال از سال پایه مورد بررسی (۲۰۱۳)، به میزان ۲۲,۲۸ درصد افزایش می‌یابد. در حالی که نرخ رشد مصرف برق بین سال‌های ۲۰۰۶ تا ۲۰۱۳ و سال‌های ۱۹۹۹ تا ۲۰۰۶ به ترتیب معادل ۲۸,۴۱ و ۷۳,۵۳ بوده است. علاوه بر این، نتایج به دست آمده از این پیش‌بینی‌ها نشان می‌دهد که مصرف برق به طور متوسط ۳,۲ درصد افزایش در هر سال خواهد داشت. در تجزیه و تحلیل انجام شده، جمعیت مهمترین عامل در مصرف برق بوده و تأثیر زیادی در افزایش میزان مصرف برق داشته است.

در مطالعه‌ای (لیو[۶] و همکاران، ۲۰۱۲)، مشتریان شرکت برق را با توجه به ماهیت آن‌ها (صنعتی، تجاری و مسکونی) و بازه‌های مصرف آن‌ها (مصرف سالانه کمتر از ۲۰۰۰ کیلووات ساعت، بیشتر از ۵۰۰۰ کیلووات ساعت یا بیشتر از ۱۸۰۰۰ کیلووات ساعت) و نوع ساختمان مشتریان خانگی (خانه‌های ویلایی، خانه‌های آپارتمانی و ساختمان‌های چند طبقه) طبقه‌بندی کرده‌اند. در این مطالعه نشان داده شده که حتی در کلاس مشتری مشابه، به دلیل ماهیت کسب و کار یا تنوع سبک زندگی مشتریان، الگوی مصرف ممکن است به میزان قابل توجهی متفاوت باشد.

---

[۱] Kalaiselvi
[۲] Abdul Samath
[۳] Kantardzic
[۴] Lytras
[۵] Error rate
[۶] Liu

در مطالعه دیگری (ناوانی[۱] و همکاران، ۲۰۱۲)، پژوهشگران با استفاده از دستیابی به الگوی مصرف برق مصرف‌کنندگان در مناطق مختلف شهری، روشی برای مدیریت بار با ظرفیت تولید برق موجود ارائه دادند. در مطالعه مذکور نشان داده شده است که مناطق مختلف، الگوی مصرف برق متفاوتی با توجه به محل و شرایط محیطی (دما، رطوبت، بارندگی و ...) و جغرافیایی آن دارد. به همین ترتیب، مصرف‌کنندگان ساکن در هر منطقه ممکن است الگوی مصرف برق مشابه و منطبق با ویژگی‌های محیطی و جغرافیایی آن منطقه داشته باشند.

در مطالعه‌ای (راتود[۲] و گارگ[۳]، ۲۰۱۶)، به بررسی مصرف برق مصرف کنندگان در هر منطقه توزیع برق و تأثیر نزدیکی ویژگی مکانی به محل مصرف کننده (منطقه) پرداخته شده است. هرگونه وابستگی ویژگی جغرافیایی به مصرف برق در قوانین وابستگی تدوین شده منعکس می‌شود. قانون وابستگی به‌کار رفته در این مطالعه، تغییرات الگوی مصرف را برای مصرف‌کنندگان ساکن در نزدیکی بزرگراه یا مزرعه (به ترتیب مصرف زیاد و کم برق) به تصویر می‌کشد. برای توصیف الگوی مصرف برق هر مصرف کننده از ویژگی‌های فضایی[۴] مانند مکان، نوع و اندازه محل سکونت، پوشش گیاهی محیط و ویژگی‌های غیرفضایی[۵] از جمله درآمد، تعداد افراد ساکن و لوازم خانگی استفاده شده است.

مدلسازی برنامه زمان‌بندی برای شبیه‌سازی انرژی مصرفی در ساختمان‌های اداری ۱۷ منطقه مختلف آب و هوایی در امریکا، پژوهش دیگری است (افخمی و همکاران، ۲۰۱۵) که به کمک ابزارهای داده‌کاوی از جمله الگوریتم‌های رگرسیون خطی، رگرسیون وزنی و رگرسیون بردار پشتیبان انجام و به ارائه مدلی برای پیش‌بینی انرژی مصرفی در یک ساختمان اداری پرداخته است. در یک پژوهش دیگر (ولازکویز[۶] و همکاران، ۲۰۱۳)، پس از تبدیل متغیرهای عددی به متغیرهای ترتیبی، یک مدل رگرسیون بین متغیر وابسته و متغیرهای مستقل بدست آورده و مدل توسعه سیستم مدیریت انرژی برای یک کارخانه پالایش بنزین به‌دست آمده است. در رابطه با ارزیابی اقتصادی شبکه توزیع نیرو (برق) با استفاده از تکنیک‌های مختلف داده‌کاوی از جمله الگوریتم ازدحام ذرات نیز مطالعه‌ای انجام شده است (ژانگ[۷]، ۲۰۱۲).

در یک مطالعه (ویلاریال[۸] و همکاران، ۲۰۱۲)، برای یکپارچه‌سازی انرژی‌های تجدیدپذیر تولید شده توسط منابع مختلف انرژی تجدیدپذیر، با توجه به متغیرهایی نظیر تعداد مطلوب و نوع برای هر جزء و محدودیت‌هایی نظیر بار مورد نیاز، مدلی را جهت حداقل کردن هزینه کل سیستم ارائه نموده است. استفاده از آمار توصیفی جهت مقایسه تکنولوژی‌های مختلف تولید و مصرف انرژی در صنعت سیمان امریکا نیز در مطالعه دیگری (مدلول[۹] و همکاران، ۲۰۱۱) به آن پرداخته شده است. همچنین با استفاده از مدل اصلاح شده بردار خطا و بررسی رابطه بین متغیرها، رابطه بین مصرف انرژی و رشد صنعت سیمان هند بررسی شده است (مندل[۱۰] و مدهسواران[۱۱]، ۲۰۱۰).

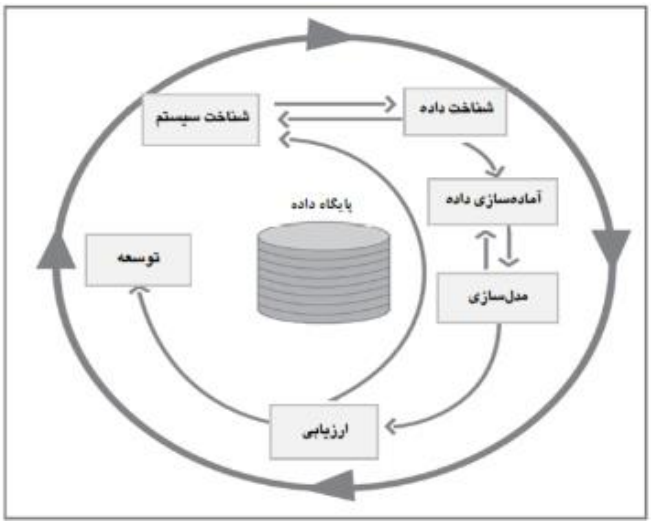
نقطه تمایز این پژوهش با سایر پژوهش‌های انجام شده، در ترکیب تکنیک خوشه‌بندی کا-میانگین و قوانین وابستگی برای بررسی مصرف انرژی واحدهای صنعتی بوده که مطابق مرور ادبیات انجام شده، تا به حال مطالعه‌ای با این ویژگی صورت نپذیرفته است. ضمناً تمرکز بر خوشه‌بندی واحدهای صنعتی بر اساس مصرف برق نیز یکی از نکات متمایز کننده این پژوهش نسبت به سایر مطالعات حوزه خوشه‌بندی می‌باشد. در بخش بعدی، ضمن اشاره به فرآیند داده‌کاوی انجام گرفته، مدل پیشنهادی تلفیقی جهت خوشه‌بندی تشریح و در ادامه، این مدل در قالب یک مطالعه موردی در رابطه با مصرف انرژی واحدهای صنعتی به کار گرفته شده است.

---

[۱] Navani
[۲] Rathod
[۳] Garg
[۴] Spatial features
[۵] Non-spatial features
[۶] Velazquez
[۷] Zhong
[۸] Villarreal
[۹] Madlool
[۱۰] Mandal
[۱۱] Madheswaran

۳- روش‌شناسی پژوهش

مزیت ا سا سی داده‌کاوی بر دیگر روش‌ها (مانند آمار)، و سعت زیاد الگوریتمی و داده‌ای آن ا ست. به عبارتی، داده‌کاوی مطابق الگوریتم‌ها و روش‌های مختلف خود تقریباً بر همه نوع داده کمی و کیفی قابلیت پیاده‌ســازی دارد. دو هدف کارکردی داده‌کاوی پیش‌بینی[۱] و توصــیف[۲] ا ست. در روش‌های پیشبینی به برآورد یک متغیر هدف با ا ستفاده از مقادیر چند متغیر دیگر پرداخته می شود ولی در نوع توصیفی که هدف اصلی این پژوهش می‌باشد، بیشترین تمرکز بر یافتن الگوها و روابط قابل توصیف در میان داده‌هاست. در این پژوهش به منظور پیاده‌سازی فرآیند داده‌کاوی، از روش ا ستاندارد داده‌کاوی میان صنعتی[۳] ا ستفاده شده ا ست. امروزه از این روش در بیشتر طرح‌های داده‌کاوی ا ستفاده می‌شود (ماربان[۴] و همکاران، ۲۰۰۹). این فرآیند دارای ۶ مرحله است که هر کدام دارای اهداف جداگانه‌ای شامل شناخت سیستم، شناخت داده، آماده‌سازی داده، مدلسازی، ارزیابی و توسعه است (اوسی بریسون[۵]، ۲۰۱۰). مراحل استاندارد این فرآیند در شکل ۱ نمایش داده شده است (چاپمن[۶] و همکاران، ۲۰۰۰). در بخش مطالعه موردی، کلیه مراحل این فرآیند تشریح خواهد شد.



شکل (۱): مراحل انجام فرآیند استاندارد داده‌کاوی

۳-۱- خوشه‌بندی

روش خوشه‌بندی[۷] یکی از زیرمجموعه‌های علم داده‌کاوی است که هدف آن، اکتشاف و پردازش پایگاه‌های داده‌ای برای استخراج دانش از آن‌هاست. خوشه‌بندی یک روش یادگیری غیرنظارتی[۸] برای دسته‌بندی داده‌ها بر اساس مشابهت‌های آن‌هاست. این روش به عنوان ابزاری توانمند جهت ا ستخراج ساختار ا صلی نهفته در مجموعه داده‌ها معرفی شده ا ست (الیویرا[۹] و پدریچ[۱۰]، ۲۰۰۷). خو شه‌بندی تکنیک مهمی ا ست که برای ت شکیل گروه‌ها یا خو شه‌های داده به کارگرفته می شود که بیانگر ویژگی م شترک کل عنا صر درون گروه ا ست (وانگ[۱۱] و همکاران، ۲۰۱۴). هر عن صر درون خو شه ن شان‌دهنده یک کلاس یا ویژگی م شترک در میان آن‌ها ا ست. الگوریتم‌های خو شه‌بندی برای

[۱] Predict
[۲] Description
[۳] CRISP Data Mining
[۴] Marban
[۵] Osei-Bryson
[۶] Chapman
[۷] Clustering
[۸] Unsupervised learning task
[۹] Oliveira
[۱۰] Pedrycz
[۱۱] Wang

انواع داده‌های مختلف از جمله عددی، دسته‌ای و چندرسانه‌ای استفاده می‌شوند (جرابا[۱]، ۲۰۰۷). خوشه‌بندی نوعی عملیات داده‌کاوی غیرمستقیم است. در اکثر روش‌های پیش‌بینی‌کننده داده‌کاوی مثل درخت تصمیم و شبکه‌های عصبی کار تحلیل با یک مجموعه آموزشی شروع شده و به کمک این مجموعه سعی می شود مدلی ایجاد شود که داده‌ها را بخش‌بندی کرده و سپس برای یک داده جدید، دسته مناسب را پیش‌بینی می‌کنند. در تحلیل خوشه‌ای هیچ دسته‌ای از قبل وجود ندارد و در واقع متغیرها به دو دسته مستقل و وابسته تقسیم نمی‌شوند. در اینجا تمرکز روی گروه‌هایی از اشیا است که به هم شبیه هستند تا با کشف این شباهت‌ها بتوان رفتار داده‌ها را بهتر شناسایی کرده و بر مبنای این شناخت بهتر تصمیم‌گیری نمود (آنتونینو[۲] و همکاران، ۲۰۱۹). در خوشه‌بندی، یک دسته داده را به چند خوشه نگاشت می‌کنیم (دورسان[۳] و کابر[۴]، ۲۰۱۶). این نکته حائز اهمیت است که این روش از سرعت رشد بسیار زیادی برخوردار بوده و در بین علوم مختلف گسترده شده است. هدف در همه روش‌ها و الگوریتم‌های خوشه‌بندی کمینه‌کردن فاصله درون خوشه‌ای و بیشینه کردن فاصله بین خوشه‌ای می‌باشد. با توجه به شرایط مختلف، انواع مختلف از خوشه‌ها نظیر خوشه‌های کروی، خوشه‌های خطی، خوشه‌های محدب و ... وجود دارد (ساتیش[۵] و یوسف[۶]، ۲۰۱۷).

۳-۲- روش خوشه‌بندی کا-میانگین[۷]

در میان الگوریتم‌های خوشه‌بندی، خوشه‌بندی کا-میانگین متداول‌ترین الگوریتم دسته‌بندی داده‌ها است. که هر رکورد واقع در مجموعه داده را فقط به یکی از خوشه‌های جدید تشکیل شده اختصاص می‌دهد. با استفاده از اندازه‌گیری فاصله یا شباهت، یک رکورد یا نقطه داده به نزدیک‌ترین خوشه (خوشه‌ای که بیشترین شباهت را به آن دارد) اختصاص می‌یابد (آنتونینو[۸] و همکاران، ۲۰۱۹). در این الگوریتم، موجودیت‌ها (رکورد داده) به نزدیک‌ترین مرکز خوشه تعلق می‌گیرند و این کار تا زمانی که هر موجودیت به نزدیک‌ترین خوشه تخصیص یابد، ادامه پیدا می‌کند (هسو[۹]، ۲۰۰۹). مجموعه داده‌هایی با N داده n بعدی xn مفروض است که هدف، تعیین تقسیم‌بندی طبیعی مجموعه داده‌ای به K خوشه است. کا-میانگین تلاش می‌کند تا مجموع مربعات تابع خوشه‌بندی *(J)* را از رابطه ۱ به حداقل برساند.

$$J = \sum_{j=1}^{k} \sum_{n \in N_j} \left\| x^n - \mu_j \right\|^2 \qquad (1)$$

که در این رابطه، μ میانگین نقاط داده‌ای در خوشه Sj است و از رابطه ۲ به‌دست می‌آید.

$$\mu_j = \frac{1}{N} \sum_{n \in S} x^n \qquad (2)$$

این دنباله با تخصیص تصادفی نقاط به K خوشه انجام می شود. سپس بردارهای میانگین μ از Nj نقطه را در هر خوشه محاسبه می‌کند. برای هر نقطه مجدداً خوشه جدیدی تعیین می‌شود که بر اساس آن بردار، نزدیک‌ترین میانگین به دست می‌آید. پس از آن، بردارهای میانگین مجدداً محاسبه می شوند. فرآیند خوشه‌بندی کا-میانگین بدین صورت قابل تعریف است که ابتدا تعداد خوشه‌ها مشخص و دسته اولیه انتخاب می‌شود و مواردی که به عضو j از دسته J که j=۱, …, k نزدیک‌ترند، تعیین می‌شوند. سپس، میانگین نمونه‌ها در هر خوشه محاسبه شده و مرکز خوشه‌ها به میانگین خوشه‌هایشان نزدیک می‌شوند. نزدیک‌ترین موارد به مرکز خوشه جدید j متعلق به خوشه j مجدداً

[۱] Djeraba
[۲] Antonino
[۳] Dursun
[۴] Caber
[۵] Satish
[۶] Yusof
[۷] K-means
[۸] Antonino
[۹] Hsu

تخصیص داده و میانگین نمونه‌ها در هر خوشه به عنوان یک مرکز خوشه جدید در نظر گرفته می‌شود. این روش آن‌قدر تکرار می‌شود تا در خوشه‌بندی تغییر بیشتری دیده نشود (لاروس[۱] و لاروس[۲]، ۲۰۱۴).

۳-۳- روش قوانین وابستگی[۳]

قوانین وابستگی[۴]، یکی از تکنیک‌های اصلی داده‌کاوی و البته احتمالاً مهمترین شکل از کشف و استخراج الگوهای محلی جالب بین متغیرها در پایگاه‌های داده بزرگ در سیستم‌های یادگیری بدون نظارت[۵] شناخته می‌شود (علیخانزاده، ۱۳۹۲). این روش در تلاش و جستجو برای یافتن ویژگی‌هایی از متغیرها ست که در بیشتر مواقع با احتمال نسبتاً بالایی با یکدیگر همراه می شوند. به عبارت دیگر، این روش به تحلیل وابستگی‌ها و مطالعه ویژگی‌ها یا خصوصیاتی می‌پردازد که با یکدیگر همراهند. قوانین وابستگی در پی یافتن ارتباط در بین آیتم‌های موجود در یک دسته از رکوردها و استخراج قواعدی به منظور کمی‌کردن ارتباط میان دو یا چند خصوصیت می‌باشد (کانگ[۶] و همکاران، ۲۰۱۷). این روش راهی برای یافتن روابط در میان آیتم‌ها یا مشخصه‌های آن‌ها که به طور همزمان در پایگاه‌های داده روی می‌دهد، می‌باشد. به صورت رسمی‌تر، قوانین وابستگی پیروی کردن پارامترها از یکدیگر را به صورت جزء به جزء توضیح می‌دهد (آخوندزاده و همکاران، ۱۳۹۵). یکی از چالش‌های اصلی در کاربرد قوانین وابستگی، چگونگی استخراج این قوانین به گونه‌ای مؤثر و کارا به ویژه در زمانی است که با پایگاه‌های داده بزرگ سر و کار داریم. در روش مرسوم برای این کار بایستی کل قواعد وابستگی ممکن از پایگاه داده استخراج شده و نرخ پشتیبانی و اطمینان هر یک محاسبه گردد (پوری، ۱۳۹۹).

قوانین وابستگی به شکل $X \rightarrow Y$ بیان می شوند که X و Y زیرمجموعه‌ای دلخواه از کل داده‌های موجود است. این سری از قوانین در جستجوی ویژگی‌های پنهانی هستند که معمولاً با یکدیگر همراه می شوند. قانون $X \rightarrow Y$ همان بیان اگر x آنگاه y را تداعی می‌کند؛ البته باید توجه داشت که ذات قوانین وابستگی، احتمالی است و بنابراین از خاصیت‌های رایجی مانند شرکت‌پذیری و توزیع‌پذیری پیروی نمی‌کنند (مارتینز[۷] و همکاران، ۲۰۱۲). برای اعتبارسنجی هر یک از این قوانین از دو معیار پشتیبان[۸] و اطمینان[۹] استفاده می شود. شاخص پشتیبان نشانگر درصدی از کل مجموعه تراکنش‌هاست که هر دو مجموعه X و Y در آن‌ها موجود باشد. اما شاخص اطمینان بیانگر میزان وابستگی مجموعه Y به X است؛ به عبارتی بیان می‌کند که چند درصد از تراکنش‌هایی که X را دارند، Y را نیز شامل می‌شوند. این شاخص‌ها با استفاده از روابط ۳ و ۴ محاسبه می‌شود.

$$Support(X) = \frac{Q\ (X)}{N} \qquad (۳)$$

$$Confidence\ (X \rightarrow Y) = \frac{Support\ (X \cup Y)}{Support\ (X)} \qquad (۴)$$

که Q(x) تعداد تعداد تراکنش‌های شامل قلم x و N تعداد کل تراکنش‌هاست.

۳-٤- تشریح مدل تلفیقی خوشه‌بندی-وابستگی

هدف اصلی این مدل کشف روابط و الگوهای پنهان موجود در پایگاه داده مورد نظر است. به همین دلیل در نتیجه استفاده از الگوریتم خوشه‌بندی و اعمال قوانین وابستگی و تلفیق این دو تکنیک، روابط و الگوهای قوی معنادار در میان داده‌های موجود شناسایی خواهد شد.

---

[۱] Larose
[۲] Larose
[۳] Association rules
[۴] در ترجمه فارسی این روش در برخی متون، از عبارات "قوانین انجمنی" و "قوانین باهم‌آیی" نیز استفاده شده است.
[۵] Undirected
[۶] Kang
[۷] Martinez
[۸] Support
[۹] Confidence

بنابراین مدل مذکور شامل دو گام است. در گام نخست بر اساس تکنینک کا-میانگین داده‌های موجود خوشه‌بندی شده و سپس در گام دوم با ا ستفاده از قوانین وابـ ستگی به ک شف الگوهای موجود در داده‌ها پرداخته خواهد شد. با ا ستفاده از خروجی این مدل، ضمن د سته‌بندی صحیح و داده‌محور، روابط معنادار وابستگی با دیگر متغیرها تشریح و در تصمیم‌گیری‌های آتی در این زمینه مورد استفاده قرار خواهد گرفت. الگوریتم خو شه‌بندی پیشنهادی با بهره‌گیری از دو روش کلا سیک کا-میانگین و قوانین وابستگی به صورت شکل ۲ ا ست. در بخش آتی، تمامی مراحل مدل پیشنهادی در قالب مطالعه موردی تشریح خواهد شد.

٤- مطالعه موردی

این مطالعه بر روی داده‌های مصرف برق واحدهای تولیدی مستقر در یکی از شهرک‌های صنعتی استان تهران به وسعت ۲۶۰ هکتار که در را ستای ساماندهی مشاغل و واحدهای تولیدی در سال ۱۳۶۹ تا سیس شده، انجام گرفته ا ست. بدین ترتیب، با توجه به داده‌های مربوط به واحدهای صنعتی و میزان مصرف برق آن‌ها در طول سال ۱۳۹۹، مدل پیشنهادی پژوهش طی مراحل زیر پیاده‌سازی شده است.

٤-١- شناخت سیستم و پایگاه داده

بر ا ساس داده‌های دریافتی از وزارت صنعت، معدن و تجارت و نیز شرکت خدماتی شهرک صنعتی مورد مطالعه، این شهرک در حال حاضر دارای ۵۳۲ واحد صنعتی فعال می‌با شد. صنایع تولیدی مجاز برای ا ستقرار در این شهرک شامل ۶ د سته " صنایع غذایی و آ شامیدنی"، "شیمیایی و پلاستیک"، "فلزی و ماشین‌سازی"، "برق و الکترونیک"، "نساجی و پوشاک" و "سلولزی" می‌باشد. بدین منظور پایگاه داده مورد مطالعه شامل اطلاعات مربوط به واحدهای صنعتی فعال و میزان مصرف برق ماهانه هر واحد در سال ۱۳۹۹ جمع‌آوری شده است.

٤-٢- انتخاب، پاکسازی و پیش پردازش داده‌ها

برای تجزیه و تحلیل داده‌های انبوه، قبل از اســتفاده از الگوریتم داده‌کاوی و در مرحله اول، عملیات پیش‌پردازش داده‌ها انجام می‌شــود (ونکاتادری[۱] و ردی[۲]، ۲۰۱۱). پیش‌پردازش داده‌ها شامل آماده سازی داده‌ها به فرم دلخواهی ا ست که پاک و عاری از هرگونه نویز ا ست. همچنین برای کاهش داده‌های انبوه واقعی به داده‌های کاربردی قابل جمع‌بندی، پیش‌پردازش داده‌ها از پردازش غیرضــروری داده‌های ناخواســته و بی‌معنی جلوگیری می‌کند (یوســف[۳] و همکاران، ۲۰۲۰). از آن‌جایی که الگوریتم کا-میانگین نســبت به داده‌های پرت، خالی و غیرنرمال حساس است، سعی بر آن شده که تا حد امکان با استفاده از روش‌های پاک‌سازی داده‌ها، از میزان داده‌های مشکل‌دار کاسته شود. برای این منظور با ا ستفاده از نرم‌افزار اک سل، مقادیر داده‌های مفقود م صرف برق ماهانه م صرف‌کنندگان، با میانگین ماه قبل و بعد از داده مفقود پر شد. همچنین تعداد ۱۸ رکورد مربوط به واحدهایی که فاقد اطلاعات م صرف برق ماهانه بودند، از مجموعه داده‌های مورد برر سی حذف و تعداد کل واحدهای صنعتی مورد مطالعه به عدد ۵۱۴ واحد ر سید (جدول ۳). پس از آماده سازی، کلیه داده‌ها وارد نرم‌افزار SPSS Modeler شده و جهت تشخیص داده‌های پرت از آزمون آنومالی[۴] ا ستفاده شد که تعداد ۵ رکورد دارای آنومالی بودند. ا ستراتژی مواجهه با این داده‌های پرت، استفاده از میانگین خانه به جای داده پرت بوده است.

٤-٣- خوشه‌بندی واحدها بر اساس مصرف انرژی

از الگوریتم خو شه‌بندی کا-میانگین در نرم‌افزار SPSS Modeler برای تـشکیل خو شه‌های م صرف‌کنندگان با ا ستفاده از داده‌های م صرف برق ماهانه در سال ۱۳۹۹ استفاده شده است. الگوریتم خوشه‌بندی کا-میانگین با استفاده از فاصله اقلیدسی[۵] بر روی ۱۲ مشخصه داده‌های

[۱] Venkatadri
[۲] Reddy
[۳] Yosef
[٤] Anomaly Detection
[٥] Euclidean distance

م صرف انرژی ماهیانه م صرف‌کنندگان اجرا می شود تا ۳ خو شه م صرف‌کنندگان بر ا ساس نوع الگوی م صرف (کم م صرف، متو سط و پرمصرف) را تشکیل دهد.

شکل (۲): گام‌های مدل پیشنهادی (مدل تلفیقی خوشه‌بندی-وابستگی)

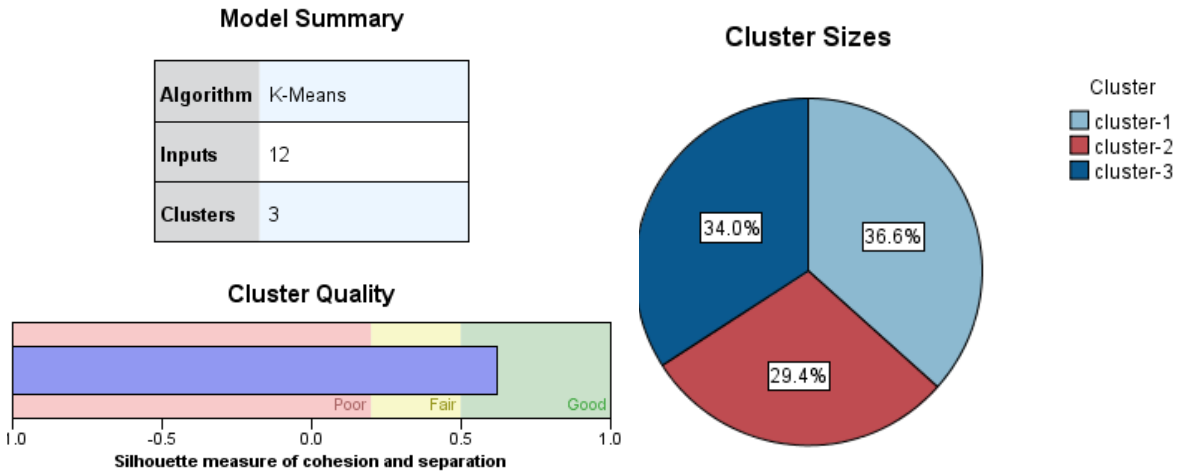
جدول (۳): تعداد و درصد واحدهای صنعتی مورد مطالعه در دسته‌های مختلف صنایع تولیدی پس از پیش پردازش

دسته تولیدی	فلزی و ماشین‌سازی	غذایی و آشامیدنی	شیمیایی و پلاستیک	برق و الکترونیک	نساجی و پوشاک	سلولزی	مجموع
تعداد واحد فعال	۲۴۱	۱۵۰	۸۲	۲۴	۲۲	۱۳	۵۳۲
تعداد واحد مورد مطالعه	۲۳۱	۱۴۸	۷۸	۲۳	۲۱	۱۳	۵۱۴
درصد واحد مورد مطالعه	۴۵ ٪	۲۹ ٪	۱۵ ٪	۵ ٪	۴ ٪	۲ ٪	۱۰۰ ٪

٤-٤- اعتبارسنجی و ارزیابی برازش مدل

برای ارزیابی کیفیت خوشه‌بندی[۱] از شاخص سیلوئت[۲] که یکی از معیارهای متداول در این زمینه می‌باشد، استفاده شده است. این شاخص دو معیار فواصل درون خوشه‌ای و برون خوشه‌ای را همزمان در نظر می‌گیرد و عددی بین ۰ و ۱ را محاسبه می‌کند. مقادیر زیر ۰,۲ نشان‌دهنده کیفیت ضعیف و بین ۰,۲ تا ۰,۵ قابل قبول است. هر چه مقدار این شاخص نزدیک به ۱ باشد، کیفیت خوشه‌بندی مناسب‌تر می‌باشد. همان‌طور که در شکل ۳ مشخص است، مدل خروجی خوشه‌بندی با ۱۲ مشخصه ورودی، شامل ۳ خوشه و شاخص سیلوئت آن ۰,۶ است که نشان‌دهنده مناسب بودن مدل خوشه‌بندی است.

همچنین ضریب اهمیت پیشبینی خوشه‌ای ۱۲ ماه سال (شکل ۴)، که نشان‌دهنده میزان اهمیت داده‌های هر ماه واحد صنعتی در نتایج خوشه‌بندی است، بین ۰,۸۴ (فروردین‌ماه) تا ۱ (شهریورماه و اردیبهشت‌ماه) است که اعدادی مناسب است که نشان‌دهنده مهم و تعیین‌کننده بودن مقادیر م صرف کلیه ماه‌های سال می‌باشد. ضمناً یکی از دلایل مهم پایین بودن ضریب اهمیت در ماه‌های فروردین و اسفند را می‌توان تعطیلی برخی از واحدهای صنعتی در روزهایی از ماه‌های مذکور به دلیل تعطیلات پایان و ابتدای سال و به تبع آن مصرف کمتر و خارج از الگوی این واحدها در این مدت ذکر کرد.

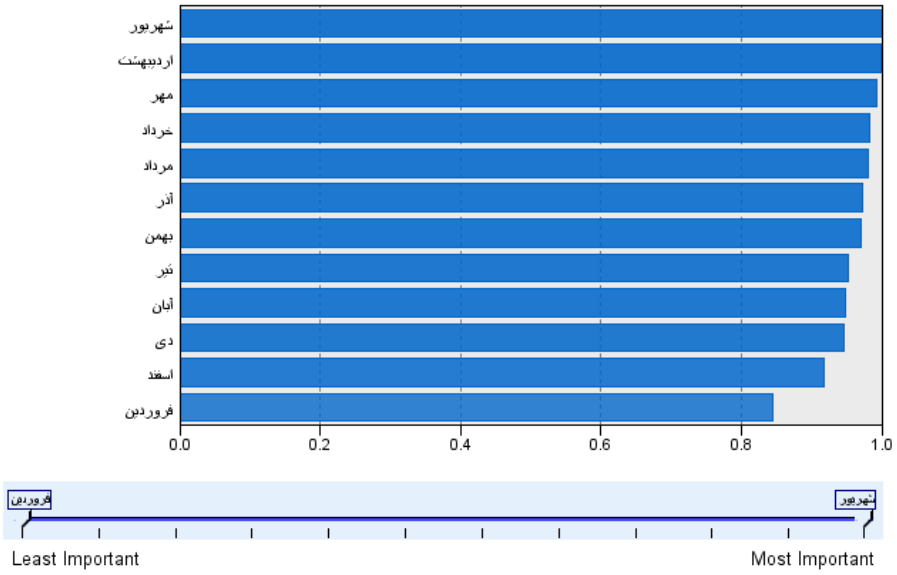


[۱] Cluster Quality
[۲] Silhouette

خوشه	برچسب	تعداد	درصد
CLUSTER ۱	کم مصرف	۱۸۸	٪ ۳۶,۶
CLUSTER ۳	متوسط	۱۷۵	٪ ۳۴,۰
CLUSTER ۲	پرمصرف	۱۵۱	٪ ۲۹,۴

Size of Smallest Cluster	151 (29.4%)
Size of Largest Cluster	188 (36.6%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.25

شکل (۳): نتایج خوشه‌بندی



شکل (۴): ضرایب اهمیت پیشبینی

٤-٥- استخراج قوانین وابستگی نوع فعالیت و مصرف برق واحدهای صنعتی

نوع فعالیت نقش مهمی در میزان مصرف انرژی، به ویژه برق داشته و تحت تأثیر بسیار زیادی از ماشین‌آلات و تجهیزات مورد استفاده است. قوانین وابستگی بهترین انتخاب برای به تصویر کشیدن روابط موجودیت‌های (رکورد داده‌های) وابسته است. الگوریتم آپریوری[۱] برای یافتن قوانین وابستگی با استفاده از شاخص‌های "پشتیبان" و "اطمینان" پیشنهاد شده است (اگراوال[۲] و اسریکانت[۳]، ۱۹۹۴). این الگوریتم به جای کنکاش در یافتن اقلام پرتکرار، به جستجو برای زیرمجموعه‌های پرتکرار در پایگاه داده می‌پردازد و چون اقلام یک زیرمجموعه پرتکرار، خود نیز همزمان پرتکرارند، پس با یک بار جستجو و خواندن پایگاه داده به راحتی می‌توان تمام اقلام تکراری را در زیرمجموعه‌های بزرگ شناسایی کرد و بدین روش نیازی به خواندن چندین مرتبه‌ای پایگاه داده توسط نرم‌افزار نخواهد بود (علیزاده و همکاران، ۱۳۹۸). به کمک قوانین وابستگی، تأثیر نوع فعالیت واحدهای صنعتی بر میزان مصرف برق مورد بررسی قرار می‌گیرد. برای دستیابی به اهداف پژوهش، وابستگی معیار نوع فعالیت با خوشه مصرف برق واحدهای صنعتی مورد بررسی قرار گرفته است. نوع فعالیت شامل یکی از ۶ دسته "صنایع غذایی و آشامیدنی"، "شیمیایی و پلاستیک"، "فلزی و ماشین‌سازی"، "برق و الکترونیک"، "نساجی و پوشاک" و "سلولزی" و خوشه‌های مصرفی شامل ۳ دسته کم‌مصرف، متوسط و پرمصرف می‌باشد. جدول ۴ نتایج حاصل را نشان می‌دهد.

جدول (۴): قوانین وابستگی استخراج شده و شاخص‌های اعتبارسنجی آن‌ها

[۱] Apriori
[۲] Agrawal
[۳] Srikant

قانون	مقدم (شرط)	تالی (نتیجه)	شاخص اطمینان	شاخص پشتیبان قانون
۱	صنایع غذایی و آشامیدنی	کم‌مصرف	٪ ۷۳	٪ ۲۲
۲	صنایع غذایی و آشامیدنی	متوسط	٪ ۴۱	٪ ۱۲
۳	صنایع غذایی و آشامیدنی	پرمصرف	٪ ۲۱	٪ ۱۲
۴	شیمیایی و پلاستیک	کم‌مصرف	٪ ۷۴	٪ ۱۲
۵	شیمیایی و پلاستیک	متوسط	٪ ۴۹	٪ ۸
۶	شیمیایی و پلاستیک	پرمصرف	٪ ۱۳	٪ ۴
۷	فلزی و ماشین‌سازی	کم‌مصرف	٪ ۴۵	٪ ۲۰
۸	فلزی و ماشین‌سازی	متوسط	٪ ۶۱	٪ ۳۲
۹	فلزی و ماشین‌سازی	پرمصرف	٪ ۲۱	٪ ۱۹
۱۰	برق و الکترونیک	کم‌مصرف	٪ ۲۱	٪ ۲
۱۱	برق و الکترونیک	متوسط	٪ ۴۷	٪ ۲
۱۲	برق و الکترونیک	پرمصرف	٪ ۶۷	٪ ۳
۱۳	نساجی و پوشاک	کم‌مصرف	٪ ۲۶	٪ ۳
۱۴	نساجی و پوشاک	متوسط	٪ ۳۲	٪ ۲
۱۵	نساجی و پوشاک	پرمصرف	٪ ۷۸	٪ ۵
۱۶	سلولزی	کم‌مصرف	٪ ۳۱	٪ ۱
۱۷	سلولزی	متوسط	٪ ۸۸	٪ ۳
۱۸	سلولزی	پرمصرف	٪ ۱۶	٪ ۱

٤-٦- تجزیه و تحلیل نتایج مدل

مدل داده‌کاوی پیشنهادی در دو فاز کار خوشه‌بندی واحدهای صنعتی را انجام داد. در فاز اول رابطه بین واحدهای صنعتی و مصرف برق با استفاده از داده‌های مصرف برق یک ساله بنا نهاده شد. الگوریتم خوشه‌بندی کا-میانگین برای تشکیل خوشه از داده‌های مصرف برق ماهیانه مصرف‌کنندگان برای بدست آوردن ۳ خوشه داده استفاده شده است. در فاز دوم، قوانین وابستگی با استفاده از الگوریتم آپریوری ایجاد شده تا قوانین وابستگی بین ۱) نوع فعالیت مصرف‌کننده و ۲) خوشه مصرف انرژی تولید و تأثیر نوع فعالیت بر مصرف برق را توسط مصرف‌کنندگان بررسی کند.

نتایج نشان می‌دهد نوع فعالیت نقش مهمی در رفتار مصرف‌کننده با توجه به میزان انرژی مصرفی ماشین‌آلات و تجهیزات دارد. نوع فعالیت هر واحد صنعتی در قالب ۶ دسته ("صنایع غذایی و آشامیدنی"، "شیمیایی و پلاستیک"، "فلزی و ماشین سازی"، "برق و الکترونیک"، "نساجی و پوشاک" و "سلولزی") مشخص شده است. بر اساس خوشه‌بندی صورت گرفته در مرحله قبل، دسته‌بندی نوع فعالیت برای تشکیل ۲ کلاس با شرایط مصرف برق مشابه (کم انرژی، انرژی‌بر) انجام شده تا تأثیر نوع فعالیت بر مصرف برق در این دوره بررسی شود. بر این اساس، دسته‌های "صنایع غذایی و آشامیدنی"، "شیمیایی و پلاستیک" و "فلزی و ماشین‌سازی" در کلاس "کم انرژی" و دسته‌های "برق و الکترونیک"، "نساجی و پوشاک" و "سلولزی" در کلاس "انرژی بر" دسته‌بندی می شوند. خلاصه نتایج در جدول ۵ آمده است.

با توجه به نتایج جدول ۶ مشاهده می‌شود در چهار ماه گرم سال (خرداد، تیر، مرداد، شهریور) که با توجه به مصرف بالای برق در کل کشور، با کمبود برق و اعمال سیاست‌های خاموشی سراسری مواجه هستیم، میانگین مصرف واحدهای خوشه کم‌مصرف حدود یک چهارم مصرف خوشه پرمصرف و حدود یک سوم مصرف خوشه متوسط است. یعنی حدود ۳۶٫۶ درصد از واحدهای صنعتی سهم بسیار کمی از مصرف برق در طی این چهار ماه را دارند و اعمال سیاست خاموشی سراسری برای تمامی واحدهای صنعتی (فارغ از دسته‌بندی مصرف) موجب ناعدالتی در اجرای قانون و عدم حمایت از واحدهایی است که به دلیل توجه به صرفه‌جویی و یا نوع فعالیت در زمره واحدهای کم مصرف بوده و ضمن تحمیل تعطیلی در ساعاتی از روز به این واحدها، انرژی چندانی هم در قبال این خاموشی از آن‌ها به شبکه تزریق نخواهد شد.

قوانین وابستگی نا شی از اجرای الگوریتم آپریوری بر روی داده‌ها با حداقل اطمینان ۶۰ در صد، منجر به ا ستخراج شش قانون وابستگی جدول ۷ شده ا ست. قوانین وابستگی به‌د ست آمده اطلاعاتی در مورد الگوی م صرف برق واحدهای مختلف را ارائه می‌دهد. این قوانین وابستگی تأثیر نوع فعالیت بر مصرف برق را توضیح می‌دهد. مصرف کمتری از برق با اطمینان ۷۴ درصد و ۷۳ درصد در واحدهای صنعتی با نوع فعالیت "شیمیایی و پلاستیک" و "صنایع غذایی و آشامیدنی" مشاهده می‌شود. در حالی که متوسط مصرف برق در واحدهایی است که نوع فعالیت آن‌ها "سلولزی" و "فلزی و ماشین‌سازی" است که به ترتیب در محدوده با اطمینان ۸۸ درصد و ۶۱ درصد است. واحدهای با نوع فعالیت "نساجی و پوشاک" و "برق و الکترونیک" نیز بیشترین مصرف برق را در مقایسه با سایر واحدها به ترتیب با اطمینان ۷۸ درصد و ۶۷ درصد دارند.

جدول (۵): کلاس انرژی واحدهای صنعتی

دسته تولیدی	فلزی و ماشین-سازی	غذایی و آشامیدنی	شیمیایی و پلاستیک	برق و الکترونیک	نساجی و پوشاک	سلولزی	کل
درصد واحد مورد مطالعه	٪ ۴۵	٪ ۲۹	٪ ۱۵	٪ ۵	٪ ۴	٪ ۲	٪ ۱۰۰
درصد واحد پرمصرف	٪ ۲۹	٪ ۲۹	٪ ۱۹	٪ ۳۷	٪ ۳۹	٪ ۲۹	٪ ۲۹,۴
درصد واحد متوسط	٪ ۳۶	٪ ۳۱	٪ ۳۶	٪ ۳۵	٪ ۲۹	٪ ۵۴	٪ ۳۴,۰
درصد واحد کم مصرف	٪ ۳۵	٪ ۴۰	٪ ۴۵	٪ ۲۸	٪ ۳۲	٪ ۱۷	٪ ۳۶,۶
کلاس انرژی	**کم انرژی**	**کم انرژی**	**کم انرژی**	**انرژی بر**	**انرژی بر**	**انرژی بر**	-

جدول (۶): میانگین مصرف خوشه‌های ۳ گانه در چهار ماه گرم سال (کیلووات ساعت)

خوشه	خرداد	تیر	مرداد	شهریور	میانگین ماه‌های گرم
پرمصرف	۱۰۱,۴۲	۱۰۸,۱۵	۱۱۰,۳۸	۱۰۱,۰۱	۱۰۵,۲۴
متوسط	۶۰,۷۳	۶۵,۲۱	۶۶,۰۰	۶۰,۹۸	۶۳,۲۳
کم مصرف	۲۴,۰۴	۲۵,۵۶	۲۵,۸۸	۲۳,۹۳	۲۴,۸۵

جدول (۷): قوانین وابستگی با حداقل اطمینان ۶۰ درصد و شاخص‌های اعتبارسنجی آن‌ها

قانون	مقدم (شرط)	تالی (نتیجه)	شاخص اطمینان	شاخص پشتیبان قانون
۱۷	سلولزی	متوسط	٪ ۸۸	٪ ۳
۱۵	نساجی و پوشاک	پرمصرف	٪ ۷۸	٪ ۵
۴	شیمیایی و پلاستیک	کم‌مصرف	٪ ۷۴	٪ ۱۲
۱	صنایع غذایی و آشامیدنی	کم‌مصرف	٪ ۷۳	٪ ۲۲
۱۲	برق و الکترونیک	پرمصرف	٪ ۶۷	٪ ۳
۸	فلزی و ماشین‌سازی	متوسط	٪ ۶۱	٪ ۳۲

۵- نتیجه‌گیری و پیشنهادهای پژوهش‌های آتی

مناطق مختلف در یک شهر الگوی م صرف برق متفاوتی دارند. در این مطالعه، از تکنیک‌های خو شه‌بندی و قوانین وابستگی برای مطالعه رفتار مصرف برق در واحدهای صنعتی فعال در یکی از شهرک‌های صنعتی استان تهران استفاده و مدل تلفیقی خوشه‌بندی-وابستگی ارائه شده ا ست. با ا ستناد بر داده‌های تجربی، م شاهده شده ا ست که م صرف برق م ستقیماً با نوع فعالیت متنا سب ا ست. این نشان می‌دهد که م صرف برق در واحدهای پرم صرف به دلیل م صرف بیشتر ما شین‌آلات، تجهیزات و و سایل برقی و الکترونیکی برای سرمایش و رو شنایی افزایش می‌یابد. م شاهدات میدانی نشان می‌دهد که طی ماه‌های گرم سال که با توجه به عدم تراز تولید و م صرف برق در ک شور منجر به اعمال سیاست‌های خاموشی اجباری در صنایع می‌شود، میانگین مصرف واحدهای خوشه پرمصرف که حدود ۳۴ درصد واحدهای صنعتی را شامل می شود، حدود ۴,۲ برابر م صرف خو شه کم‌م صرف و حدود ۱,۷ برابر م صرف خو شه متو سط ا ست. به عبارت دیگر با شنا سایی

هوشمندانه واحدهای پرمصرف و اعمال سیاست‌های عادلانه در خاموشی اجباری، می‌توان علاوه بر تشویق واحدهای صنعتی به بهینه‌سازی مصرف انرژی، از ایجاد خسارت به کسب و کار آن‌ها با توقف‌های اجباری تولید ممانعت کرد. این مدل داده‌کاوی با توجه به الگوریتم تشریح شده، برای هر مجموعه واحدهای صنفی دیگر نیز قابل تعمیم و استفاده است. مدل ارائه شده قادر است واحدهای صنفی را از طریق بررسی رفتار مصـرف برق آن‌ها با اسـتفاده از مجموعه قوانین وابسـتگی متمایز کند. رویکرد نوآورانه این مدل قادر به کنترل حجم زیادی از داده‌ها برای برنامه‌ریزی مناطق مختلف صنفی با هدف بهینه‌سازی در مصرف برق است. قانون وابستگی، خلاصه تغییرات الگوی مصرف را برای مصـرف‌کنندگان با توجه به نوع فعالیت و کلاس انرژی به تصـویر می‌کشـد. دانش حاصـل شـده در رابطه با نوع فعالیت مصـرف‌کننده، به پیش‌بینی رفتار مصرف برق آن‌ها کمک می‌کند. در پژوهش‌های آتی می‌توان از این مدل و تلفیق سایر الگوهای داده‌کاوی برای پیش‌بینی مصـرف انرژی در برنامه‌ریزی شـهری نیز اسـتفاده نمود. با توجه به توانمندی این مدل در تجزیه و تحلیل رفتار مصـرف سـایر بخش‌های مصرف کننده فارغ از نوع انرژی مصرفی، مدل می تواند در شناسایی رفتار مصرف سایر انرژی‌های مصرفی نیز مفید واقع گردد.

منابع

[۱] آخوندزاده نوقابی، الهام. دانشـــمندی، مهدی. مینایی بیدگلی، بهروز. (۱۳۹۵). ارایه رویکردی برای کاهش هزینه‌های کیفیت در اندازه‌گیری پارامترهای شیمیایی با استفاده از تکنیک داده‌کاوی. مهندسی صنایع و مدیریت شریف. ۳۲: ۹-۱.

[۲] آذر، عادل. خدیور، آمنه. (۱۳۹۸). کاربرد تحلیل آماری چندمتغیره در مدیریت. نگاه دانش. تهران.

[۳] پوری، کوروش. (۱۳۹۹). خو شه‌بندی بازار تو سط تکنیک‌های داده‌کاوی با رویکرد ترکیبی K-means و قواعد انجمنی. اولین کنفرانس بین‌المللی چالش‌ها و راهکارهای نوین در مهندسی صنایع، مدیریت و حسابداری. تهران.

[۴] ترازنامه انرژی سال ۱۳۹۷. (۱۳۹۹). دفتر برنامه‌ریزی و اقتصاد کلان برق و انرژی، تهران.

[۵] علیخانزاده، امیر. (۱۳۹۲). داده‌کاوی. نشر علوم رایانه. تهران.

[۶] علیزاده، سمیه. تیمورپور، بابک. غضنفری، مهدی. (۱۳۹۸). داده‌کاوی و کشف دانش. دانشگاه علم و صنعت ایران.

[۷] قره‌خانی، محسن. ابوالقاسمی، مریم. (۱۳۹۰). تازه‌های جهان بیمه، ۱۴: ۱۵۸. ۲۲-۵.

[۸] Agrawal R., Srikant R. (۱۹۹٤). Fast algorithms for mining association rules. ۲۰th international conference of very large databases. Santiago, Chile.

[۹] Antonino A., Massimo F., La Rosa V. (۲۰۱۹). Data mining: classification and prediction. Encyclopedia of Bioinformatics and Computional Biology. ۳۸٤-٤۰۲.

[۱۰] Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. (۲۰۰۰). CRISP-DM ۱,۰ Step-by-step Data mining guide. SPSS Inc.

[۱۱] Djeraba C. (۲۰۰۷). Data mining from multimedia. International journal of Parallel, Emergent and Distributed Systems. ٤۰٥-٤۰٦.

[۱۲] Dursun A., Caber M., (۲۰۱٦). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. Tourism Management Perspectives. ۱۸. ۱٥۳-۱٦۰.

[۱۳] Hsu, C. H. (۲۰۰۹). Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry. Expert Systems with Applications. ۳٦. ٤۱۸٥-٤۱۹۱.

[۱٤] Kalaiselvi, K., Abdul Samath J. (۲۰۱۹). Survey on electricity consumption using data mining techniques. Int. J. Advanced Networking and Applictions. ۱۰(٤). ۳۹۷٤-۳۹۸۰.

[۱٥] Kang, S., Kim, E., Shim J., Cho S., Chang W., Kim, J. (۲۰۱۷). Mining the relationship between production and customer service data for failure analysis of industrial products. Computers and Industrial Engineering. ۱۰٦. ۱۳۷-۱٤٦.

[۱٦] Kantardzic M. (۲۰۱۹). Data mining: concepts, models, methods and algorithms. John wiley & sons.

[۱۷] Karimtabar, N., Pasban, S., Alipour, S. (۲۰۱۵). Analysis and predicting electricity energy consumption using data mining techniques. ۲nd International Conference on Pattern Recognition and Image Analysis (IPRIA ۲۰۱۵). ۱۱-۱۲.
[۱۸] Larose D., Larose C., (۲۰۱٤). Discovering knowledge in data: an introduction to data mining. John wiley & Sons.
[۱۹] Liu H., Yao Z., Eklund T., Back B. (۲۰۱۲). Electricity consumption time series profiling: A data mining application in energy industry. Industrial Conference on Data Mining. ٥٢-٥٦.
[۲۰] Lytras M., Visvizi A., Zhang X., Aljohani N. R. (۲۰۲۰). Cognitive computing, Big data analytics and data driven industrial marketing. Industrial Marketing Management. ۹۰. ٦٦٣-٦٦٦.
[۲۱] Madlool N.A., Saidur R., Hossain M.S., Rahim N.A. (۲۰۱۱). A critical review on energy use and savings in the cement industries. Renewable and Sustainable Energy Reviews. (۱٥)٤. ۲۰٤۲-۲۰٦۰.
[۲۲] Mandal S. K., Madheswaran S. (۲۰۱۰). Causality between energy consumption and output growth in the Indian cement industry: An application of the panel vector error correction model (VECM). Energy Policy. ۳۸(۱۱). ٦٥٦۰-٦٥٦٥.
[۲۳] Marban O., Segovia J., Menasalvas E., Fernandez-Biazan C. (۲۰۰۹). Toward data mining engineering: A software engineering approach. Information Systems. ۳٤(۱). ۸۷-۱۰۷.
[۲٤] Martinez-de-Pison, F.J., Sanz, A., Martinez-de-Pison, E., Jimenez, E. and Conti, D. (۲۰۱۲). Mining association rules from time series to explain failures in a hot-dip galvanizing steel line. Journal of Computers & Industrial Engneering. ٦۳(۱). ۲۲-۳٦.
[۲٥] Navani J. P., Sharma N. K., Sapra S. (۲۰۱۲). Technical and non-technical losses in power system and its economic consequence in Indian economy. International Journal of Electronics & Computer Science Engineering. (۱)۷٥۷-۷٦۱.
[۲٦] Oliveira J. V., Pedrycz W. (۲۰۰۷). Advances in fuzzy clustering and its applications. New York: Wiley.
[۲۷] Osei-Bryson K. (۲۰۱۰). Towards supporting expert evaluation of clustering results using a data mining process model. Information Sciences. ۱۸(۳). ٤۱٤-٤۳۱.
[۲۸] Rathod R., Garg R. (۲۰۱٦). Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. Electrical Power & Energy Systems. ۷۸. ۳٦۸-۳۷٤.
[۲۹] Satish L., Yusof N. (۲۰۱۷). A review: Big data analytics for enhanced customer experiences with crowd sourcing. Procedia Computer Science. ۱۱٦. ۲۷٤-۲۸۳.
[۳۰] Velazquez D., Gonzalez-Falcon R., Perez-Lombard L., Gallego L. M., Monedero I., Biscarri F. (۲۰۱۳). Development of an energy management system for a naphtha reforming plant: A data mining approach. Energy Conversion and Management. ٦۷. ۲۱۷-۲۲٥.
[۳۱] Venkatadri M., Reddy L. (۲۰۱۱). A review on data mining from past to the future. International Journal of Computer Applications. (۱٥)۹. ۱۹-۲۲.
[۳۲] Villarreal C. M., Lopez N., Espiritu J. F. (۲۰۱۲). Using the Monkey Algorithm for Hybrid Power Systems Optimization. Procedia Computer Science. ۱۲. ۳٤٤-۳٤۹.
[۳۳] Wang Z., Tu L., Guo Z. (۲۰۱٤). Analysis of user behaviors by mining large network data sets. Future generation of computer systems. ۳۷. ٤۲۹-٤۳۷.
[۳٤] Yoseph F., Malim N. H., Heikkila M., Brezulianu A., German O., Rostam N. P. (۲۰۲۰). The impact of big data market segmentation using mining and clustering techniques. Journal of Intelligent & Fuzzy Systems. ۳۸(٥). ٦۱٥۹-٦۱۷۳.
[۳٥] Zhong J. (۲۰۱۲). Application of Graph Data Mining Method in Distribution Network Economic Evaluation. Physics Procedia. ۳۳. ٦۱۲-٦۱۸.