

طراحی مدل داده کاوی تلفیقی خوشبندی-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

عظیم‌الله زارعی

استاد مدیریت، دانشگاه سمنان

A_Zarei@semnan.ac.ir

چکیده

تاریخ دریافت:
۱۴۰۱/۳/۸

تاریخ پذیرش:
۱۴۰۱/۱۲/۱۶

کلمات کلیدی:

داده کاوی

خوشبندی

تجزیه و تحلیل سلسله مراتبی

الگوریتم کا-میانگین

قوانین وابستگی

بهینه‌سازی مصرف انرژی

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 داده کاوی 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، مدل تلفیقی جهت بررسی رفتار مصرف برق با استفاده از تکنیک‌های خوشبندی 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 خوش پر-صرف که حدود ۳۴ درصد واحدهای صنعتی مورد مطالعه را شامل می‌شود، حدود ۴,۲ برابر مصرف خوش کم‌صرف و حدود ۱,۷ برابر مصرف خوش متوسط است. با بکارگیری مدل پیشنهادی در این پژوهش ضمن شناسایی واحدهای پر-صرف 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 پیشافت 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 می‌شود. 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 مصرف بهینه آن در اشكال مختلف اقداماتی را برنامه‌ریزی و اجرا می‌نمایند. با بررسی 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 انرژی بر ناشی می‌شود. 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 می‌شود 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 ۲۶ 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 گونه‌ای 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 می‌رساند. در بخش صنعت نیز استان اصفهان با ۱۴۴۰,۸,۹ گیگاوات ساعت بیشترین میزان مصرف برق را به خود اختصاص داده است. در جدول ۲، درصد رشد سالانه مصرف حامل‌های انرژی در بخش صنعت به تفکیک حامل‌های انرژی طی سال‌های ۱۳۹۷ تا ۱۳۸۹ 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

درصد سهم	۳۴	۲۲,۸	۱۶,۶	۱۴,۵	۱,۹	۰,۲	حمل و نقل	سایر مصارف	کشاورزی	تجاری و عمومی	صنعت	بخش
----------	----	------	------	------	-----	-----	-----------	------------	---------	---------------	------	-----

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

کل مصرف انرژی	برق	زغال سنگ	گاز طبیعی	فرآورده‌های نفتی
۱,۹۸	۸,۰۴	-۷۱,۶۲	۱۷,۷۶	۱۱,۶۳
۹,۳۸	۷,۸۶	-۲,۴۶	۶,۱۵	۲,۹۹
۵,۹۷	۴,۱۲	۰	۴۳,۳۲	۵,۹۳
۰,۹۱	۹,۶۵	-۰,۰۷	۵,۷۶	۱۸,۳۳
۵,۳۸	۴,۰۷	-۲,۱۷	۷,۳۹	-۲,۶۸
-۴,۹۶	۱,۷۶	۲۱,۲۷	۳,۹۵	۵,۹۷
۶,۱۳	۴,۵۶	۴۳,۳۲	۶,۷۵	۲,۶۸
۴,۱۹	۷,۱۰	۴۳,۳۲	-۱۳,۱۴	-۳,۲۹
۵,۷۲	۲,۰۷	۲۱,۲۷	۳,۹۵	۱۳۹۴
۹,۳۸	۱,۹۸	۱۳۹۰	۱۳۸۹	۱۳۹۷
۱,۹۸	۱۳۹۱	۱۳۹۲	۱۳۹۳	۱۳۹۴
-۳۴,۶۵	۶,۹۸	-۱۱,۳۷	۰,۶۳	-۲,۶۸
-۱۰,۴۸	-	-	-	-

۱-۲-۱ داده کاوی

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). امروزه داده کاوی به عنوان یکی از ابزارهای بسیار مهم مدیران جهت شناخت و ضعیت دقیق سازمان و همچنین کمک در اتخاذ تصمیمات 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 داده کاوی می‌تواند توصیف یا پیش‌بینی است و می‌توان گفت داده کاوی 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، داده کاوی 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 داده کاوی و غفلت از نتایج کاربردی و مطابق با واقعیت‌های محیط مورد بررسی است. این امر ناچیز از آن است که به دلیل پیچیدگی‌های ذی‌سی روش‌های داده کاوی، بیشتر وقت و انرژی پژوهشگران بر حل مسائل تکنیکی متمرکز می‌شود. به علاوه کاربرد روش‌های گوناگون داده کاوی ممکن است به نتایج متفاوتی نیز بینجامد و لازم است پژوهشگران نتایج روش‌های مختلف داده کاوی را مقایسه و آن‌ها را در کنار شرایط محیط مطالعه، به کار گیرند (لیتراس^۴ و همکاران، ۲۰۲۰).

۲- مرور ادبیات و تحقیقات پیشین

بر اساس تجزیه و تحلیل‌های تحت نظرارت 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 ۲۰۱۳ تا ۲۰۰۶ 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 (صرف 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 م‌شایعه، 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 مصرف کنندگان در مناطق مختلف شهری، رو شی برای مدیریت بار با ظرفیت تولید برق موجود را 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، مصرف کنندگان 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 تزدیکی 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 به کار رفته در این مطالعه، تغییرات الگوی مصرف را برای مصرف کنندگان 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 (اخخی و همکاران، ۲۰۱۵) که به کمک ابزارهای داده کاوی 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 (ولادکویز^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 به دست آمده است. در رابطه با ارزیابی اقتصادی شبکه توزیع نیرو (برق) با استفاده از تکنیک‌های مختلف داده کاوی از جمله الگوریتم ازدحام ذرات نیز مطالعه‌ای انجام شده است (زانگ^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 (مدلو^۹ و همکاران، ۲۰۱۱) به آن پرداخته شده است. همچنین با استفاده از مدل اصلاح شده بردار خطأ و بررسی رابطه بین متغیرها، رابطه بین مصرف انرژی و رشد صنعت سیمان هند بررسی شده است (مدل^{۱۰} و مدهسوaran^{۱۱}).

نقشه تمایز این پژوهش با سایر پژوهش‌های انجام شده، در ترکیب تکنیک خوشبندی 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 خوشبندی واحدهای صنعتی بر اساس مصرف برق نیز یکی از نکات تمایز کننده این پژوهش نسبت به سایر مطالعات حوزه خوشبندی می‌باشد. در بخش بعدی، ضمن اشاره به فرآیند داده کاوی انجام گرفته، مدل پیشنهادی تلفیقی جهت خوشبندی تشریح و در ادامه، این مدل در قالب یک مطالعه موردي 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

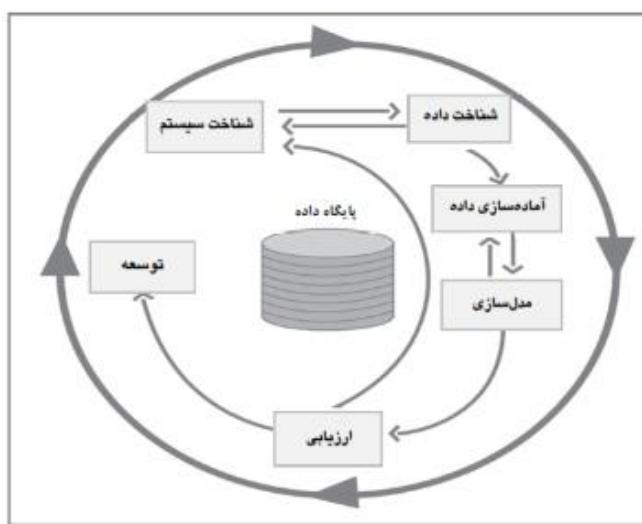
^۹ Madlool

^{۱۰} Mandal

^{۱۱} Madheswaran

۳- روش‌شناسی پژوهش

مزیت اساسی داده‌کاوی بر دیگر روش‌ها (مانند آمار)، و سمعت زیاد الگوریتمی و داده‌ای آن است. به عبارتی، داده‌کاوی مطابق الگوریتم‌ها و روش‌های مختلف خود تقریباً بر همه نوع داده کمی و کیفی قابلیت پیاده‌سازی دارد. دو هدف کارکردی داده‌کاوی پیش‌بینی^۱ و توصیف^۲ است. در روش‌های پیش‌بینی به برآورد یک متغیر هدف با استفاده از مقادیر چند متغیر دیگر پرداخته می‌شود 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 استاندارد داده‌کاوی میان صنعتی^۳ استفاده شده است. امروزه از این روش در بیشتر طرح‌های داده‌کاوی استفاده می‌شود (ماربان^۴ و همکاران، ۲۰۰۹). این فرآیند دارای ۶ مرحله است که هر کدام دارای اهداف جدایهای شامل شناخت سیستم، شناخت داده، آماده‌سازی داده، مدل‌سازی، ارزیابی و توسعه است (اوی برسون^۵، ۲۰۱۰). مراحل استاندارد این فرآیند در شکل ۱ نمایش داده شده است (چاپمن^۶ و همکاران، ۲۰۰۰). در بخش مطالعه موردي، کلیه مراحل این فرآیند تشریح خواهد شد.



شکل (۱): مراحل انجام فرآیند استاندارد داده‌کاوی

۱-۳- خوشبندی

روش خوشبندی^۷ یکی از زیرمجموعه‌های علم داده‌کاوی است که هدف آن، اکتشاف و پردازش پایگاه‌های داده‌ای برای استخراج دانش از آن‌هاست. خوشبندی یک روش یادگیری غیرنظراتی^۸ برای دسته‌بندی داده‌ها بر اساس مشابهت‌های آن‌هاست. این روش به عنوان ابزاری توانمند جهت استخراج ساختار اصلی نهفته در مجموعه داده‌ها معرفی شده است (الیویرا^۹ و پدریچ^{۱۰}، ۲۰۰۷). خوشبندی تکنیک مهمی است که برای تشکیل گروه‌ها یا خوشه‌های داده به کارگرفته می‌شود که بیانگر ویژگی مشترک کل عناصر درون گروه است (وانگ^{۱۱} و همکاران، ۲۰۱۴). هر عنصر درون خوشه نشان‌دهنده یک کلاس یا ویژگی مشترک در میان آن‌ها است. الگوریتم‌های خوشبندی 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 چند رسانه‌ای استفاده می‌شوند (جرابا^۱، ۲۰۰۷). خوشبندی نوعی عملیات داده کاوی غیرمستقیم است. در اکثر روش‌های پیش‌بینی کننده داده کاوی مثل درخت تصمیم و شبکه‌های عصبی کار تحلیل با یک مجموعه آموزشی شروع شده و به کمک این مجموعه سعی می‌شود مدلی ایجاد شود که داده‌ها را بخش‌بندی کرده و سپس برای یک داده جدید، دسته مناسب را پیش‌بینی می‌کنند. در تحلیل خوشبندی 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 خوشبندی، یک دسته داده را به چند خوشبندی 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 خوشبندی کمینه کردن فاصله درون خوشبندی و بیشینه کردن فاصله بین خوشبندی می‌باشد. با توجه به شرایط مختلف، انواع مختلف از خوشبندی نظیر خوشبندی کروی، خوشبندی خطی، خوشبندی محدب و وجود دارد (ساتیش^۵ و یوسف^۶، ۲۰۱۷).

۲-۳- روش خوشبندی کا-میانگین^۷

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

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

که در این رابطه، μ میانگین نقاط داده‌ای در خوشبندی S_j است و از رابطه ۲ به دست می‌آید.

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

این دنباله با تخصیص تصادفی نقاط به K خوشبندی انجام می‌شود. سپس بردارهای میانگین μ از N نقطه را در هر خوشبندی محاسبه می‌کند. برای هر نقطه مجددًا خوشبندی جدیدی تعیین می‌شود که بر اساس آن بردار، نزدیک‌ترین میانگین به دست می‌آید. پس از آن، بردارهای میانگین مجددًا محاسبه می‌شوند. فرآیند خوشبندی کا-میانگین بدین صورت قابل تعریف است که ابتدا تعداد خوشبندی مشخص و دسته اولیه انتخاب می‌شود و مواردی که به عضو j از دسته J که $k = 1, \dots, j$ نزدیک‌ترند، تعیین می‌شوند. سپس، میانگین نمونه‌ها در هر خوشبندی محاسبه شده و مرکز خوشبندی که میانگین خوشبندی ایشان نزدیک می‌شوند. نزدیک‌ترین موارد به مرکز خوشبندی جدید j متعلق به خوشبندی j مجددًا

^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 خوش به عنوان یک مرکز خوش جدید در نظر گرفته می‌شود. این روش آنقدر تکرار می‌شود تا در خوشبندی تغییر بیشتری دیده نشود (لاروس^۱ و لاروس^۲، ۲۰۱۴).

۳-۳- روش قوانین وابستگی^۳

قوانین وابستگی^۳، 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^۴ شناخته می‌شود (علیخانزاده، ۱۳۹۲). این روش در تلاش و جستجو برای یافتن ویژگی‌هایی از متغیرهاست که در بیشتر مواقع با احتمال نسبتاً بالایی با یکدیگر همراه می‌شوند. به عبارت دیگر، این روش به تحلیل وابستگی‌ها و مطالعه ویژگی‌ها یا خصوصیاتی می‌پردازد که با یکدیگر همراهند. قوانین وابستگی در بیان ارتباط در بین آیتم‌های موجود در یک دسته از رکوردها و استخراج قواعدی به منظور کمی‌کردن ارتباط میان دو یا چند خصوصیت می‌باشد (کانگ^۵ 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 (آنوندزاده 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 مر سوم 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 می‌شوند 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 می‌شوند. قانون $Y \rightarrow X$ 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^۶ و همکاران، ۲۰۱۲). برای اعتبار سنجی هر یک از این قوانین از دو معیار پشتیبان^۷ و اطمینان^۸ استفاده می‌شود. 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.

$$Support(X) = \frac{Q(X)}{N} \quad (3)$$

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \quad (4)$$

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

۳-۴- تشریح مدل تلفیقی خوشبندی-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 خوشبندی 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 تکینیک کا-میانگین داده‌های موجود خوشبندی شده و سپس در گام دوم با استفاده از قوانین وابستگی به کشف الگوهای موجود در داده‌ها پرداخته خواهد شد. با استفاده از خروجی این مدل، ضمن دسته‌بندی صحیح و داده‌محور، روابط معنادار وابستگی با دیگر متغیرها تشریح و در تصمیم‌گیری‌های آتی در این زمینه مورد استفاده قرار خواهد گرفت. الگوریتم خوشبندی پیشنهادی با بهره‌گیری از دو روش کلاسیک کا-میانگین و قوانین وابستگی به صورت شکل ۲ است. در بخش آتی، تمامی مراحل مدل پیشنهادی در قالب مطالعه موردی تشریح خواهد شد.

۴- مطالعه موردی

این مطالعه بر روی داده‌های مصرف برق واحدهای تولیدی مستقر در یکی از شهرک‌های صنعتی استان تهران به وسعت ۲۶ هکتار که در راستای ساماندهی مشاغل و واحدهای تولیدی در سال ۱۳۶۹ تاسیس شده، انجام گرفته است. 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.

۴-۱- شناخت سیستم و پایگاه داده

بر اساس 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 می‌باشد. صنایع تولیدی مجاز برای استقرار در این شهرک شامل ۶۰ دسته "صنایع غذایی و آشامیدنی"، "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.

۴-۲- انتخاب، پاکسازی و پیش‌پردازش داده‌ها

برای تجزیه و تحلیل داده‌های انبو، قبل از استفاده از الگوریتم داده‌کاوی و در مرحله اول، عملیات پیش‌پردازش داده‌ها انجام می‌شود (ونکاتادری^۱ و ردی^۲، ۲۰۱۱). پیش‌پردازش داده‌ها شامل آماده سازی داده‌ها به فرم دلخواهی است که پاک و عاری از هرگونه نویز است. همچنین برای کاهش داده‌های انبو واقعی به داده‌های کاربردی قابل جمع‌بندی، پیش‌پردازش داده‌ها از پردازش غیرضروری داده‌های ناخواسته و بی‌معنی جلوگیری می‌کند (یوسف^۳ و همکاران، ۲۰۲۰). از آن جایی که الگوریتم کا-میانگین نسبت 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 استفاده از نرم‌افزار اکسل، مقادیر داده‌های مفقود مصرف برق ماهانه مصرف‌کنندگان، با میانگین ماه قبل و بعد از داده مفقود پر شد. همچنین تعداد ۱۸ رکورد مربوط به واحدهایی که فقد اطلاعات مصرف برق ماهانه بودند، از مجموعه داده‌های مورد بررسی حذف و تعداد کل واحدهای صنعتی مورد مطالعه به عدد ۵۱۴ واحد ر رسید (جدول ۳). پس از آماده سازی، کلیه داده‌ها وارد نرم‌افزار SPSS Modeler شده و جهت تشخیص داده‌های پرت از آزمون آنومالی^۴ استفاده شد که تعداد ۵ رکورد دارای آنومالی بودند. استراتژی مواجهه با این داده‌های پرت، استفاده از میانگین خانه به جای داده پرت بوده است.

۴-۳- خوشبندی واحدهای بر اساس مصرف انرژی

از الگوریتم خوشبندی کا-میانگین در نرم‌افزار SPSS Modeler برای تشكیل خوشبندی مصرف کنندگان با استفاده از داده‌های مصرف برق ماهانه در سال ۱۳۹۹ استفاده شده است. الگوریتم خوشبندی کا-میانگین با استفاده از فاصله اقلیدسی^۵ بر روی ۱۲ مشخصه داده‌های

^۱ Venkatadri

^۲ Reddy

^۳ Yosef

^۴ Anomaly Detection

^۵ Euclidean distance

مصرف انرژی ماهیانه مصرف کنندگان اجرا می شود تا ۳ خوش مصرف (کم مصرف، متوسط و پرمصرف) را تشکیل دهد.



شکل (۲): گام‌های مدل پیشنهادی (مدل تلقیقی خوشبندی-وابستگی)

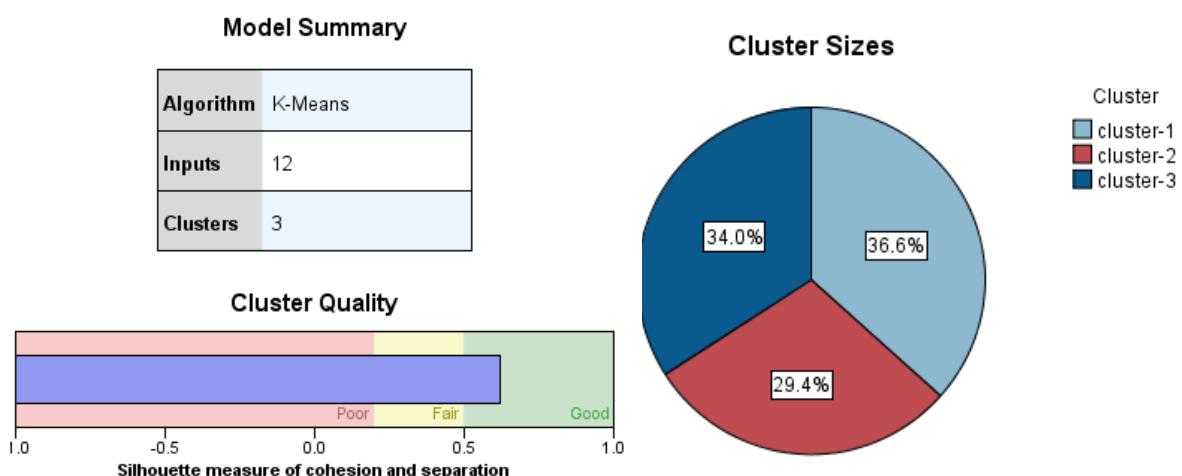
جدول (۳): تعداد و درصد واحدهای صنعتی مورد مطالعه در دسته‌های مختلف صنایع تولیدی پس از پیش پردازش

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

۴-۴- اعتبارسنجی و ارزیابی برآذش مدل

برای ارزیابی کیفیت خوشبندی^۱ از شاخص سیلوئت^۲ که یکی از معیارهای متداول در این زمینه می‌باشد، استفاده شده است. این شاخص دو معیار فواصل درون خوشبندی و برون خوشبندی را هم‌زمان در نظر می‌گیرد و عددی بین ۰ و ۱ را محاسبه می‌کند. مقادیر زیر ۰,۲ نشان‌دهنده کیفیت ضعیف و بین ۰,۵ تا ۰,۵ قابل قبول است. هرچه مقدار این شاخص نزدیک به ۱ باشد، کیفیت خوشبندی مناسب‌تر می‌باشد. همان‌طور که در شکل ۳ مشخص است، مدل خروجی خوشبندی با ۱۲ مشخصه ورودی، شامل ۳ خوش و شاخص سیلوئت آن ۰,۶۰ است که نشان‌دهنده مناسب بودن مدل خوشبندی است.

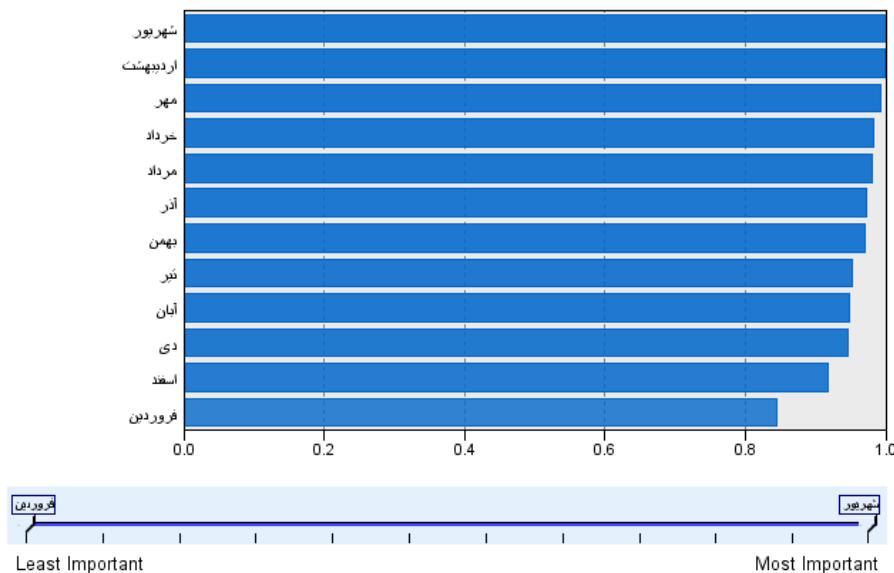
همچنین ضریب اهمیت پیشینی خوشبندی ۱۲ ماه سال (شکل ۴)، که نشان‌دهنده میزان اهمیت داده‌های هر ماه واحد صنعتی در نتایج خوشبندی است، بین ۰,۸۴ (فروردين‌ماه) تا ۱ (شهریور‌ماه و اردیبهشت‌ماه) است که اعدادی مناسب است که نشان‌دهنده مهم و تعیین‌کننده بودن مقادیر مصرف کلیه ماههای سال می‌باشد. ضمناً یکی از دلایل 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

درصد	تعداد	برچسب	خوشه	
% ۳۶.۶	۱۸۸	کم مصرف	CLUSTER ۱	Size of Smallest Cluster 151 (29.4%)
% ۳۴.۰	۱۷۵	متوسط	CLUSTER ۳	Size of Largest Cluster 188 (36.6%)
% ۲۹.۴	۱۵۱	پرمصرف	CLUSTER ۲	Ratio of Sizes: Largest Cluster to Smallest Cluster 1.25

شکل (۳): نتایج خوشبندی



شکل (۴): ضرایب اهمیت پیشیبینی

۴-۵- 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 پر تکرار، به جستجو برای زیرمجموعه‌های پر تکرار در پایگاه داده می‌پردازد و چون اقلام یک زیرمجموعه پر تکرار، خود نیز همزمان پر تکرارند، 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 "صناعی غذایی و آشامیدنی"، "شیمیایی و پلاستیک"، "فلزی و ماشین‌سازی"، "برق و الکترونیک"، "نساجی و پوشاک" و "سلولزی" و خوشه‌های مصرفی شامل ۳ دسته کم مصرف، 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	کم مصرف	۷۳ %	۲۲ %	
۲	صنایع غذایی و آشامیدنی	متوسط	۴۱ %	۱۲ %	
۳	صنایع غذایی و آشامیدنی	پرمصرف	۲۱ %	۱۲ %	
۴	شیمیابی و پلاستیک	کم مصرف	۷۴ %	۱۲ %	
۵	شیمیابی و پلاستیک	متوسط	۴۹ %	۸ %	
۶	شیمیابی و پلاستیک	پرمصرف	۱۳ %	۴ %	
۷	فلزی و ماشین سازی	کم مصرف	۴۵ %	۲۰ %	
۸	فلزی و ماشین سازی	متوسط	۶۱ %	۳۲ %	
۹	فلزی و ماشین سازی	پرمصرف	۲۱ %	۱۹ %	
۱۰	برق و الکترونیک	کم مصرف	۲۱ %	۲ %	
۱۱	برق و الکترونیک	متوسط	۴۷ %	۲ %	
۱۲	برق و الکترونیک	پرمصرف	۶۷ %	۳ %	
۱۳	نساجی و پوشک	کم مصرف	۲۶ %	۳ %	
۱۴	نساجی و پوشک	متوسط	۳۲ %	۲ %	
۱۵	نساجی و پوشک	پرمصرف	۷۸ %	۵ %	
۱۶	سلولزی	کم مصرف	۳۱ %	۱ %	
۱۷	سلولزی	متوسط	۸۸ %	۳ %	
۱۸	سلولزی	پرمصرف	۱۶ %	۱ %	

۴-۶- تجزیه و تحلیل نتایج مدل

مدل داده کاوی پیشنهادی در دو فاز کار خوشبندی 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 خوشبندی کامیانگین برای تشکیل خوش از دادههای مصرف برق ماهیانه مصرف کنندگان برای بدست آوردن ۳ خوش داده استفاده شده است. در فاز دوم، قوانین وابستگی با استفاده از الگوریتم آپریوری ایجاد شده تا قوانین وابستگی بین ۱) نوع فعالیت مصرف کننده و ۲) خوش مصرف انرژی تولید و تأثیر نوع فعالیت بر مصرف برق را توسط مصرف کنندگان بررسی کند.

نتایج نشان می‌دهد نوع فعالیت نقش مهمی در رفتار مصرف کننده با توجه به میزان انرژی مصرفی ماشین‌آلات و تجهیزات دارد. نوع فعالیت هر واحد صنعتی در قالب ۶ دسته ("صنایع غذایی و آشامیدنی"، "شیمیابی و پلاستیک"، "فلزی و ماشین سازی"، "برق و الکترونیک"، "نساجی و پوشک" و "سلولزی") مشخص شده است. بر اساس خوشبندی صورت گرفته در مرحله قبل، دسته‌بندی نوع فعالیت برای تشکیل ۲ کلاس با شرایط مصرف برق مشابه (کم انرژی، انرژی بر) انجام شده تا تأثیر نوع فعالیت بر مصرف برق در این دوره بررسی شود. بر این اساس، دسته‌های "صنایع غذایی و آشامیدنی"، "شیمیابی و پلاستیک" و "فلزی و ماشین سازی" در کلاس "کم انرژی" و دسته‌های "برق و الکترونیک"، "نساجی و پوشک" و "سلولزی" در کلاس "انرژی بر" دسته‌بندی می‌شوند. 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 خوش کم مصرف حدود یک چهارم مصرف خوش پرمصرف و حدود یک سوم مصرف خوش 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 ناعادالتی در اجرای قانون و عدم حمایت از واحدهایی است که به دلیل توجه به صرفه‌جویی و یا نوع فعالیت در زمرة واحدهای کم مصرف بوده و ضمن تحمیل تعطیلی در ساعتی از روز به این واحدها، انرژی چندانی هم در قبال این خاموشی از آن‌ها به شبکه تزریق نخواهد شد.

قوانين وابستگی نا شی از اجرای الگوریتم آپریوری بر روی داده ها با حداقل اطمینان ۶۰ درصد، منجر به استخراج شش قانون وابستگی جدول ۷ شده است. قوانین وابستگی بدست آمده اطلاعاتی در مورد الگوی مصرف برق واحدهای مختلف را ارائه می دهد. این قوانین وابستگی تأثیر نوع فعالیت بر مصرف برق را توضیح می دهد. مصرف کمتری از برق با اطمینان ۷۴ درصد و ۷۳ درصد در واحدهای صنعتی با نوع فعالیت "شیمیابی و پلاستیک" و "صناعی غذایی و آشامیدنی" مشاهده می شود. در حالی که متوسط مصرف برق در واحدهایی است که نوع فعالیت آنها "سلولزی" و "فلزی و ماشین سازی" است که به ترتیب در محدوده با اطمینان ۸۸ درصد و ۶۱ درصد است. واحدهای با نوع فعالیت "نساجی و پوشک" 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	شیمیابی و پلاستیک	برق و الکترونیک	سلولزی	نساجی و پوشک	کل
درصد واحد مورد مطالعه	۴۵%	۲۹%	۱۵%	۵%	۴%	۲%	۱۰۰%
درصد واحد پر مصرف	٪ 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 خوشهای ۳ گانه در چهار ماه گرم سال (کیلووات ساعت)

کم مصرف	متوسط	پر مصرف
۶۰,۷۳	۶۶,۰۰	۱۱۰,۳۸
۲۴,۰۴	۲۵,۵۶	۱۰۱,۴۲

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

قانون	مقدم (شرط)	تالی (نتیجه)	شاخص اطمینان	شاخص پشتیبان قانون	خوشه	تیر	مرداد	شهریور	میانگین ماههای گرم
۱۷	سلولزی	متوسط	۸۸%	۳%	۱۰۱,۰۱	۱۰۵,۳۴			
۱۵	نساجی و پوشک	پر مصرف	۷۸%	۵%	۶۰,۹۸	۶۳,۲۳			
۴	شیمیابی و پلاستیک	کم مصرف	۷۴%	۱۲%	۶۵,۲۱	۶۰,۷۳			
۱	صناعی غذایی و آشامیدنی	کم مصرف	۷۳%	۲۲%	۲۵,۸۸	۲۳,۹۳	۲۴,۸۵		
۱۲	برق و الکترونیک	پر مصرف	۶۷%	۳%	۲۵,۵۶				
۸	فلزی و ماشین سازی	متوسط	۶۱%	۳۲%	۲۴,۰۴				

۵- نتیجه گیری و پیشنهادهای پژوهش های آتی

مناطق مختلف در یک شهر الگوی مصرف برق متفاوتی دارند. در این مطالعه، از تکنیک های خوب شنبه و قوانین وابستگی برای مطالعه رفتار مصرف برق در واحدهای صنعتی فعال در یکی از شهر که های صنعتی استان تهران استفاده و مدل تلفیقی خوب شنبه وابستگی ارائه شده است. با استناد بر داده های تجربی، مشاهده شده است که مصرف برق مستقیماً با نوع فعالیت مناسب است. این نشان می دهد که مصرف برق در واحدهای پر مصرف به دلیل مصرف بیشتر ماشین آلات، تجهیزات و وسایل برقی و الکترونیکی برای سرمایش و روشنایی افزایش می یابد. مشاهدات میدانی نشان می دهد که طی ماههای گرم سال که با توجه به عدم تراز تولید و مصرف برق در کشور منجر به اعمال سیاست های خاموشی اجباری در صنایع می شود، میانگین مصرف واحدهای خوشه پر مصرف که حدود ۳۴ درصد واحدهای صنعتی را شامل می شود، حدود ۴,۲ برابر مصرف خوشه کم مصرف و حدود ۱,۷ برابر مصرف خوشه متوسط است. به عبارت دیگر با شناشی

هوشمندانه واحدهای پرصرف 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 مصرف برق آن‌ها با استفاده از مجموعه قوانین وابستگی 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 تقریرات الگوی مصرف را برای مصرف‌کنندگان با توجه به نوع فعالیت و کلاس انرژی به تصویر می‌کشد. دانش حاصل شده در رابطه با نوع فعالیت مصرف‌کننده، به پیش‌بینی رفتار مصرف برق آن‌ها کمک می‌کند. در پژوهش‌های آتی می‌توان از این مدل و تلفیق سایر الگوهای داده‌کاوی برای پیش‌بینی مصرف انرژی در برنامه‌ریزی شهری نیز استفاده نمود. با توجه به توانمندی این مدل در تجزیه و تحلیل رفتار مصرف سایر بخش‌های مصرف کننده فارغ از نوع انرژی مصرفی، مدل می‌تواند در شناسایی رفتار مصرف سایر انرژی‌های مصرفی نیز مفید واقع گردد.

منابع

- [۱] آخوندزاده نوqابی، الهام. دانشمندی، 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: ۱-۹.
- [۲] آذر، عادل. خدیور، آمنه. (۱۳۹۸). کاربرد تحلیل آماری چندمتغیره در مدیریت. نگاه دانش. تهران.
- [۳] پوری، کوروش. (۱۳۹۹). خوشه‌بندی 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 ریانه. 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: ۱۵۸-۵۲۲.
- [۸] Agrawal R., Srikant R. (۱۹۹۴). 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 Computational Biology. ۳۸۴-۴۰۲.
- [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. ۴۰۵-۴۰۶.
- [۱۲] Dursun A., Caber M., (۲۰۱۶). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. Tourism Management Perspectives. ۱۸. ۱۵۳-۱۶۰.
- [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. ۳۶. ۴۱۸۰-۴۱۹۱.
- [۱۴] Kalaiselvi, K., Abdul Samath J. (۲۰۱۹). Survey on electricity consumption using data mining techniques. Int. J. Advanced Networking and Applications. ۱۰(۴). ۳۹۷۴-۳۹۸۰.
- [۱۵] 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. ۱۰۶. ۱۳۷-۱۴۶.
- [۱۶] Kantardzic M. (۲۰۱۹). Data mining: concepts, models, methods and algorithms. John Wiley & Sons.

- [۱۷] Karimtabar, N., Pasban, S., Alipour, S. (۲۰۱۰). Analysis and predicting electricity energy consumption using data mining techniques. ۴nd International Conference on Pattern Recognition and Image Analysis (IPRIA ۲۰۱۰). ۱۱-۱۲.
- [۱۸] Larose D., Larose C., (۲۰۱۴). 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. ۵۲-۵۶.
- [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. ۶۶۳-۶۶۶.
- [۲۱] Madloli 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. (۱۰)۴. ۲۰۴۲-۲۰۶۰.
- [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(11). ۶۰۶۰-۶۰۶۰.
- [۲۳] Marban O., Segovia J., Menasalvas E., Fernandez-Biazan C. (۲۰۰۹). Toward data mining engineering: A software engineering approach. Information Systems. ۳۴(1). ۸۷-۱۰۷.
- [۲۴] 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 Engineering. ۶۳(1). ۲۲-۳۶.
- [۲۵] 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)۷۰۷-۷۱۱.
- [۲۶] 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). ۴۱۴-۴۳۱.
- [۲۸] Rathod R., Garg R. (۲۰۱۱). Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. Electrical Power & Energy Systems. ۷۸. ۳۶۸-۳۷۴.
- [۲۹] Satish L., Yusof N. (۲۰۱۷). A review: Big data analytics for enhanced customer experiences with crowd sourcing. Procedia Computer Science. ۱۱۶. ۲۷۴-۲۸۳.
- [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. ۷۷. ۲۱۷-۲۲۰.
- [۳۱] Venkatadri M., Reddy L. (۲۰۱۱). A review on data mining from past to the future. International Journal of Computer Applications. (۱۰)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. ۳۴۴-۳۴۹.
- [۳۳] Wang Z., Tu L., Guo Z. (۲۰۱۴). Analysis of user behaviors by mining large network data sets. Future generation of computer systems. ۳۷. ۴۲۹-۴۳۷.
- [۳۴] 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(۵). ۶۱۰۹-۶۱۷۳.
- [۳۵] Zhong J. (۲۰۱۲). Application of Graph Data Mining Method in Distribution Network Economic Evaluation. Physics Procedia. ۳۳. ۶۱۲-۶۱۸.