

ارائه چارچوب پیشنهادی جامع جهت تجزیه و تحلیل داده‌ها

در سیستم مدیریت انرژی سازمانی

کامران رضایی^۱، حسین قدمیان^۲، عبدالرحمن حائری^۳

چکیده

آنالیز داده‌ها جزء لاینفک هر سیستم مدیریتی در یک سازمان، از جمله سیستم مدیریت انرژی می‌باشد. بوسیله آنالیز داده-هاست که سازمان می‌تواند بر اساس واقعیات به تصمیم‌گیری و سیاست گذاری در مورد سیستم مدیریت انرژی سازمان بپردازد. علاوه بر این، ارزیابی میزان اثربخشی و موفقیت / شکست برنامه‌های بهبود کارایی انرژی تنها از طریق جمع‌آوری داده-های مرتبط و تجزیه و تحلیل آنها میسر می‌باشد. در این مقاله تلاش شده است تا یک رویه منطقی جهت آنالیز داده‌های مربوط به یک سیستم مدیریت انرژی ارائه شود. رویه پیشنهادی بر اساس الگوریتم CRISP-DM بیان شده است. این رویه شامل شش مرحله می‌باشد. مراحل درک فضای کسب و کار، شناسایی داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، پایش و ارزشیابی و اجرا می‌باشد.

کلمات کلیدی: سیستم مدیریت انرژی، مدیریت داده‌ها، آنالیز داده‌ها، الگوریتم CRISP-DM

مقدمه

در طول سالیان گذشته حجم زیادی از داده‌ها در پایگاه‌های اطلاعاتی سازمان‌ها انباشته شده است. نتیجه این انباشتگی این است که سازمانها با حجم عظیمی از داده‌ها مواجه هستند. در دنیای امروز داده‌ها و اطلاعات حاصل از آنها به منزله یکی از دارایی‌های سازمان جهت ایجاد ارزش افزوده و تولید ثروت به حساب می‌آیند. اما در فضای کنونی سازمان‌ها در کسب دانش از

^۱ عضو هیات علمی دانشکده فنی دانشگاه تهران، مدیر عامل شرکت توف نورد ایران، Email: energy@tuvnordiran.com

^۲ عضو هیات علمی دانشگاه، مدیر عامل شرکت کار و نیرو، Email: energy@karoniroo.com

^۳ * دانشجوی دکتری مهندسی صنایع، دانشکده فنی دانشگاه تهران، Email: ahaeri@ut.ac.ir

داده‌ها با چالش‌های عمده‌ای مواجه هستند. در واقع امروزه سازمان‌ها از حجم بسیار زیاد و روزافزون داده‌های خود استفاده چندانی نکرده و در بسیاری از موارد تنها به ذخیره سازی آنها می‌پردازند. در حالیکه یکی از منابع ایجاد ارزش افزوده داده‌ها می‌باشند. برای بهره‌گیری درست از داده‌های سازمانی مراحل زیر به صورت عمومی و صرف نظر از نوع داده‌ها و هدف از پردازش سه مرحله کلی زیر ضروری می‌باشد:

- جمع آوری و ذخیره سازی داده‌ها
- آنالیز و استخراج اطلاعات از داده‌های بدست آمده
- استفاده از اطلاعات جهت تصمیم‌گیری (Decision Making) موثر و بهبود مستمر

در نتیجه می‌توان گفت که برای تصمیم‌گیری موثر و صحیح در جنبه‌های مختلف سازمان از جمله سیستم مدیریت انرژی، چاره‌ای جز بهره‌گیری موثر از داده‌ها و اطلاعات حاصل از آنها نمی‌باشد. در این مقاله به بررسی یک رویه پیشنهادی جهت تعیین اهداف، جمع‌آوری و آنالیز داده‌های سیستم مدیریت انرژی در سازمان‌ها پرداخته می‌شود.

الگوریتم CRISP-DM

یکی از رویکردهای آنالیز و تحلیل داده‌ها، رویکرد داده‌کاوی (Data Mining) می‌باشد که به کشف دانش از داده می‌پردازد. برای داده‌کاوی الگوریتم‌ها و رویه‌های مختلفی توسعه داده شده است. یکی از این الگوریتم‌ها که به خوبی می‌تواند تحلیل گران را در استخراج دانش از داده‌ها یاری رساند، الگوریتم استاندارد CRISP-DM می‌باشد. از آنجاییکه متدولوژی مذکور مستقل از نوع داده‌ها بوده و به خوبی می‌تواند جهت آنالیز داده‌ها به صورت عام مورد استفاده قرار گیرد، در این مقاله از آن به عنوان مبنا برای آنالیز داده‌ها در سیستم مدیریت انرژی استفاده می‌شود. این الگوریتم در شش مرحله قابل بیان می‌باشد که در ادامه به تشریح مراحل آن و فعالیت‌های مرتبط با آنالیز داده‌ها در سیستم مدیریت انرژی در هر مرحله می‌پردازد.

۱) درک فضای کسب و کار (Business understanding)

این مرحله در حکم شالوده و اساس آنالیز داده‌ها می‌باشد. در این مرحله اهداف کسب و کار در زمینه آنالیز داده‌ها مشخص شده و مورد بحث و بررسی قرار می‌گیرند. این فاز شامل فعالیت‌های زیر می‌باشد:

- تعیین اهداف تجاری
- ارزیابی موقعیت

- تعیین اهداف آنالیز داده‌ها
 - تهیه برنامه و رویه‌های مربوط به آنالیز داده‌ها
- در یک سیستم مدیریت انرژی اهداف زیر به عنوان اهداف کسب و کار مطرح می‌باشند:
- کاهش هزینه‌های عملیاتی
 - کاهش قیمت تمام شده محصول که می‌تواند پیامدهای زیر را به دنبال داشته باشد:
 - افزایش رضایت مشتریان
 - افزایش سهم بازار
 - افزایش توان رقابتی
 - ارزیابی و تعیین اولویت بین فرصت‌های صرفه‌جویی انرژی

۲) شناسایی داده‌ها (Data understanding)

داده‌ها در واقع مواد خام برای آنالیز به حساب می‌آیند. در این فاز به شناسایی منابع داده‌ها و مشخصات آنها پرداخته می‌شود.

این مرحله شامل فعالیت‌های زیر می‌باشد:

- جمع‌آوری داده‌های اولیه
- تشریح و توصیف داده‌ها
- کاوش داده‌ها
- اعتبارسنجی کیفیت داده‌ها

پیشنهاد می‌شود که در این مرحله یک ارزیابی از سیستم اطلاعاتی انرژی سازمان نیز انجام شود. این ارزیابی شامل مراحل زیر

می‌باشد:

- تعیین داده‌های مورد نیاز جهت اهداف مشخص شده در فاز اول
- ارزیابی پایگاه داده‌های مرتبط با انرژی در سازمان
- تعیین نواقص پایگاه داده‌های موجود
- ارائه راهکار، دستورالعمل و روش اجرایی برای جمع‌آوری داده‌های اطلاعات مورد نیاز در آینده

داده‌های مورد استفاده در یک سیستم مدیریت انرژی بسیار متنوع می‌باشند. برخی از این داده‌ها عبارتند از:

- ساعات کاری/ اشغال فضا (Operating Hours/Occupancy): در ساختمان‌های اداری، مدارس، بیمارستان‌ها، فروشگاه‌ها، انبارها و ... مصرف انرژی در درجه اول به هوای سیستم‌های گرم کننده و سرد کننده بستگی داشته و در درجه دوم به الگوهای مصرف از قبیل ساعات کاری (operating hours) و سکونت (occupancy) بستگی دارد.
- سطح تولید (Production Level): مصرف حامل‌های انرژی در تجهیزات صنعتی و تولیدی ارتباط مستقیمی با سطح تولید یا معیارهای خروجی تجهیزات از قبیل واحد، تن و ... دارد.
- دمای هوا (Weather Degree): مصرف حامل‌های انرژی با دما ارتباط تنگاتنگی دارد. به همین علت است که مصرف حامل‌های انرژی در فصول سرد و گرم تفاوت معناداری با هم دارد.
- داده‌های مربوط به سه فاز ممیزی انرژی در
- قبوض انرژی (Energy Bills): که نوعاً شامل قبض برق و گاز می‌باشد.
- داده‌های موقعیت جغرافیایی
- نقشه ساختمان
- ساعات انجام عملیات
- لیست تجهیزات
- داده‌های حاصل از بازدید از ساختمان، کارخانه یا سایت که شامل داده‌های مربوط به موارد زیر می‌باشد:
 - بررسی پوشش‌ها و نقاط مرزی ساختمان
 - دیگ بخار و سیستم توزیع بخار
 - سیستم HVAC
 - سیستم ذخیره و پشتیبانی برق
 - چراغ‌ها، پنجره‌ها و سطوح صیقلی منعکس کننده
 - سیستم توزیع آب گرم
 - سیستم فشرده‌سازی و توزیع هوا
 - موتورها
 - فرآیندهای تولیدی [1]
- داده‌های حاصل از اجرای توصیه‌های ممیزی مانند میزان مصرف حامل‌های انرژی
- میزان صرفه‌جویی تخمینی حاصل از فرصت‌های صرفه‌جویی انرژی (Energy Conservation Opportunity=ECO)

۳) آماده‌سازی داده‌ها (Data preparation)

پس از جمع‌آوری داده‌ها بایستی به آماده‌سازی و به عبارتی پیش‌پردازش داده‌ها برای آنالیز و تجزیه و تحلیل داده‌ها پرداخت. این مرحله شامل فعالیت‌های زیر می‌باشد:

- انتخاب داده‌ها
- پاکسازی داده‌ها
- ساختاردهی به داده‌ها
- یکپارچه سازی داده‌ها
- تنظیم فرمت داده‌ها
- تعریف و یا انتخاب شاخص‌های مناسب (در صورت لزوم)

فازهای دوم و سوم معمولاً وقت زیادی را به خود اختصاص می‌دهند. البته این زمان بستگی به زیر ساخت اطلاعاتی سازمان مورد نظر دارد.

این فاز از اهمیت زیادی برخوردار است. یکی از فعالیت‌هایی که باید در طول این فاز انجام شود، این است که داده‌ها استاندارد (نرمالایز) شوند. برای درک بهتر موضوع یک مثال ذکر می‌شود. فرض کنید که یک شرکت دارای دو کارخانه به شماره‌های ۱ و ۲ می‌باشد که میزان مصرف حامل‌های انرژی در کارخانه ۱، یک و نیم برابر میزان مصرف در کارخانه شماره ۲ می‌باشد. آیا این امر به معنای میزان اتلاف بیشتر و وجود فرصت‌های بیشتر برای صرفه‌جویی انرژی در سالن شماره ۱ می‌باشد؟ پاسخ منفی است. چرا که میزان مصرف حامل‌های انرژی به فاکتورهای متعددی از قبیل میزان تولید، دما و میزان فضا بستگی دارد. از این رو باید شاخص‌هایی مناسب و استاندارد تعریف و انتخاب شوند تا بتوان بر اساس آنها به درستی به تجزیه و تحلیل داده‌ها و تصمیم‌گیری بر اساس نتایج حاصل از آنها پرداخت. دو نمونه از این شاخص‌ها عبارتند از:

• شاخص بهره‌برداری انرژی (Energy Utilization Index=EUI)

مقدار این شاخص برابر با مقدار انرژی مورد استفاده در هر فوت مربع در هر سال می‌باشد.

• شاخص هزینه انرژی (Energy Cost Index=ECI)

مقدار این شاخص برابر با هزینه انرژی مورد استفاده در هر سال در هر فوت مربع در فضای مورد نظر می‌باشد.

۴ مدل سازی (Modeling)

این فاز مرحله اصلی انجام آنالیز روی داده‌ها می‌باشد. در این مرحله از روش‌ها و تکنیک‌های مختلفی جهت جهت آنالیز داده‌ها و استخراج دانش از آنها به کار گرفته می‌شوند. این مرحله شامل فعالیت‌های زیر می‌باشد:

- انتخاب تکنیک‌های ساخت مدل
- طراحی آزمایش
- ساخت مدل
- ارزیابی مدل

برخی از تکنیک‌های مورد استفاده و موارد کاربرد آنها در جدول ۱ ذکر شده‌اند.

ردیف	تکنیک / رویکرد مورد استفاده	کاربرد
۱	سری‌های زمانی	پیش‌بینی مصرف حامل‌های انرژی (مثلاً اینکه با تغییر دمای هوا یا میزان تولید چه تغییری در میزان مصرف انرژی حاصل می‌شود) [4]
۲	آنالیز فرآیند آماری (SPC)	شناسایی علت نوسان در مصرف انرژی در یک فرآیند
۳	رگرسیون و شبکه عصبی	محاسبه میزان بودجه انرژی، تخمین تابع / معادله (مثلاً توسعه مدل محاسبه میزان مصرف حامل‌های انرژی بر اساس میزان تولید، فضا و دما) [2]
۴	تحلیل پوششی داده‌ها (DEA)	ارزیابی عملکرد و رتبه‌بندی واحدها/ فرآیندهای مختلف سازمان از جهت مصرف حامل‌های انرژی
۵	خوشه‌بندی (Clustering) و کلاس‌بندی (Classification)	شناسایی الگوها در بین داده‌های عملکردی مربوط به مصرف حامل‌های انرژی [3]

شناسایی علل موفقیت / شکست برنامه‌های افزایش کارایی انرژی	کشف قواعد وابستگی (Association Rule Mining)	۶
بهینه‌سازی میزان مصرف حامل‌های انرژی در سازمان/ بخش / فرآیند	روش‌های بهینه‌سازی (Optimization Methods)	۷
اولویت‌بندی فرصت‌های صرفه‌جویی انرژی	آنالیزهای اقتصادی از قبیل محاسبه ارزش فعلی، نرخ و دوره بازگشت سرمایه	۸
		۹

جدول ۱: برخی از ابزارهای مورد استفاده در آنالیز داده‌ها و کاربرد آنها

۵) پایش و ارزشیابی (Evaluation)

پس از انتخاب مدل بایستی در این مرحله به بررسی و ارزیابی این موضوع پرداخت که آیا نتایج آنالیز داده‌ها ما را در رسیدن به اهداف کسب و کار یاری می‌کنند؟ در این مرحله دانش بدست آمده در مرحله چهارم مورد تجزیه و تحلیل قرار خواهد گرفت تا میزان سودمندی و کاربرد آن مشخص شود. به عنوان مثال در مورد مدل‌های پیش‌بینی کننده باید دقت مدل روی داده‌های جدید تست شود و در صورت تایید نتایج حاصل از آن می‌تواند به کار گرفته شود. این مرحله شامل فعالیت‌های زیر می‌باشد:

- ارزیابی نتایج
- مرور فرآیند آنالیز
- تعیین گام‌های بعدی

۶ اجرا (Deployment)

در این مرحله سود حاصل از سرمایه‌گذاری انجام شده در مراحل قبلی به دست می‌آید. تمرکز این فاز روی یکپارچه‌سازی دانش در فرآیندهای کاری کسب و کار به گونه‌ای است که مسائل اساسی کسب و کار حل شوند. این مرحله شامل فعالیت‌های زیر می‌باشد:

- اجرای برنامه
- نظارت و نگهداری
- آماده‌سازی گزارش نهایی
- مرور پروژه

ذکر این نکته ضروری است که هر سازمان بسته به زیرساخت‌های خود و سیستم مدیریت انرژی آن ممکن است بر روی تعدادی از مراحل و زیرمرحله‌ها تمرکز بیشتری داشته باشد. در نتیجه مرور فرآیند بالا مشخص شد که آنالیز داده‌ها تنها مجموعه‌ای از تکنیک‌ها و روش‌های محاسباتی و ریاضی نیست. بلکه یک فرآیند منسجم و یکپارچه بوده که با شناسایی اهداف و نیازهای فضای کسب و کار (فاز اول) آغاز می‌شود. سپس به تناسب نیازهای مورد نظر، داده‌ها جمع‌آوری و آماده‌سازی (فازهای دوم و سوم) می‌شوند. سپس آنالیز روی داده‌ها بسته به کاربردهای مورد نظر (فاز چهارم) انجام می‌گردد. در فاز پنجم نتایج آنالیزهای انجام شده مورد ارزیابی و اعتبارسنجی قرار می‌گیرند. در نهایت در فاز ششم، دانش و نتایج بدست آمده در فرآیندهای کسب و کار جهت بهبود و نیل به اهداف فاز اول به کار گرفته می‌شود.

نتیجه‌گیری و پیشنهاد تحقیقات آتی

در این مقاله به بررسی یک رویه پیشنهادی جهت آنالیز داده‌ها در سیستم مدیریت انرژی در سازمان پرداخته شد. رویه پیشنهادی از الگوریتم داده‌کاوی CRISP-DM که جهت کشف دانش از داده کاربرد دارد، نشأت گرفته است. رویه پیشنهادی شامل فعالیت‌های اصلی زیر بود:

۱) درک فضای کاری و شناسایی اهداف مرتبط با انرژی در سازمان

۲) شناسایی و جمع‌آوری داده‌های مورد نیاز

۳) آماده‌سازی داده‌ها

۴) مدل‌سازی

۵) پایش و ارزیابی

۶) پیاده‌سازی و اجرا

در پایان لازم به ذکر است که هر یک از کاربردهای شناسایی شده در جدول ۱ خود به تنهایی می‌تواند موضوع یک مقاله مستقل تحقیقاتی در آینده باشد.

مراجع

- 1- Barney L. Capehart, Wayne C. Turner, William J. Kennedy, Guide to energy management, Guide to Energy Management, Taylor & Francis Group, 2005
- 2- Wayne C. Turner, Steve Doty, Energy management handbook, sixth edition, Taylor & Francis Group, 2006
- 3- Albert Thumann, P.E., C.E.M. William J. Younger, C.E.M., Handbook of energy audits, seventh edition, Taylor & Francis Group, 2007
- 4- Dale R. Patrick, Stephen W. Fardo, Ray E. Richardson, Steven R. Patrick, Energy Conservation Guidebook, Second Edition, Taylor & Francis Group, 2006