

## به کارگیری تئوری مجموعه‌های راف در تحلیل نتایج ممیزی انرژی

تورج کریمی<sup>۱</sup>

**چکیده:** هدف از این تحقیق، استخراج مدل قوانین مربوط به داده‌های حاصل از ممیزی انرژی ساختمان‌ها به کمک تئوری راف است. تئوری راف دارای الگوریتم‌های قدرتمندی است که امکان تحلیل داده‌ها را فراهم می‌کند. ابزارهای استفاده‌شده در این تئوری می‌تواند با ارزش‌های نادقیق و داده‌های غیرقطعی کار کرده و واقعیت‌های پنهان در داده‌ها را کشف کند. از آنجاکه قسمتی از گزارش‌های ممیزی انرژی ساختمان مربوط به سنجش سطح آسایش ساکنان ساختمان از نظر وضعیت سرمایش و گرمایش است و بخش دیگر داده‌ها مربوط به بررسی‌های فنی ساختمان است، در این تحقیق یک مشخصه تصمیم مربوط به سطح آسایش ساکنان و یازده مشخصه موقعیتی مربوط به جنبه‌های فنی ساختمان تحلیل شده و استنتاج قوانین به کمک نرم‌افزار ROSETTA صورت گرفته است. با توجه به الگوریتم‌های مختلف تکمیل داده‌ها، گسسته کردن مقادیر و تولید بی‌زائده، براساس شرایط تحقیق چهار مدل از قوانین ساخته شده و به روش اعتبارسنجی موازی نتایج مدل‌ها ارزیابی شده است. در مجموع بهترین مدل با ۱۴ قانون و دقت پیش‌بینی ۸/۹۹ درصد انتخاب شده است. نتایج این مدل نشان داد که مهم‌ترین مشخصه فنی هر ساختمان «متراز فضای کنترل‌نشده» است و با اطلاع از مقدار این مشخصه می‌توان سطح آسایش کارکنان در ساختمان را با دقت زیادی پیش‌بینی کرد.

واژه‌های کلیدی: استنتاج قوانین، مجموعه‌های راف، ممیزی انرژی ساختمان، ROSETTA.

۱. دکتری تولید و عملیات، دانشکده مدیریت، دانشگاه تهران، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۳/۱۲/۲۶

تاریخ پذیرش نهایی مقاله: ۱۳۹۴/۰۲/۲۱

E-mail: tkarimi@ut.ac.ir

### مقدمه

پس از بحران انرژی دهه ۱۹۷۰، مطالعات بسیاری در خصوص مصرف انرژی از جانب رشته‌های گوناگون نظیر اقتصاد، تکنولوژی، جامعه‌شناسی، رفتارشناسی و نظایر آن صورت گرفته است. لوتزن‌هیسر که از دهه ۱۹۹۰ موضوع انرژی و رفتار را بررسی کرده، معتقد است بررسی مصرف انرژی بخش‌های پویای بسیاری دارد که به راحتی تفسیرپذیر نیستند و نمی‌توان تنها بعد تکنولوژی یا انسانی آن را دید یا تنها با رویکرد سیاستگذاری‌های اقتصادی آن را بررسی کرد (لوتزن‌هیسر، ۲۰۰۸).

استیفنسون، چارچوب فرهنگ‌های انرژی<sup>۱</sup> یا ECF را معرفی کرد. این چارچوب بیان می‌دارد که رفتارهای مصرف انرژی را با بررسی و دقت در تعامل میان هنجارهای شناختی<sup>۲</sup> (نظیر اعتقادات و دانسته‌ها)، فرهنگ مادی<sup>۳</sup> (نظیر فناوری‌ها) و عادات‌های انرژی<sup>۴</sup> (نظیر فعالیت‌ها و فرایندها) می‌توان شناخت که این سه جزء رفتار به شدت بر هم اثرگذارند (استیفنسون و همکاران، ۲۰۱۰).

اگرچه در فرایند ممیزی انرژی ساختمان‌ها بررسی جنبه‌های فنی ساختمان اهمیت زیادی دارد، به دلایل توضیح داده شده در بالا بررسی برخی از جنبه‌های غیرفنی نظیر سطح آسایش ساکنان نیز در این ممیزی‌ها گنجانده شده است. در این تحقیق سعی شده بین عوامل مربوط به بعد فنی و بعد انسانی در مصرف انرژی ارتباط برقرار شود.

با اینکه شدت مصرف انرژی ساختمان به‌ازای هر متر مربع در ایران دو تا سه برابر استانداردهای جهانی است، شرایط آسایشی حرارتی در داخل ساختمان‌ها اغلب با انتقاد مواجه شده و اکثر ساختمان‌های کشور، از ضوابط فنی شناخته شده برای جلوگیری از هدر رفتن انرژی بی‌بهره‌اند. در این تحقیق به کمک تئوری راف و نرم‌افزار ROSETTA، مدل قوانین برای بررسی سطح آسایش کارکنان ساکن در ساختمان‌های اداری استخراج و اعتبارسنجی شده است. در این مدل شاخص‌های فنی مرتبط با مصرف انرژی در ساختمان‌های اداری به‌عنوان مشخصه‌های موقعیتی، و شاخص آسایش کارکنان از نظر گرمایش و سرمایش به‌عنوان مشخصه تصمیم در نظر گرفته شده است. این تحقیق هم از جهت نوع توجه به موضوع انرژی و هم از لحاظ روش بررسی دارای نوآوری است؛ چراکه در آن، اطلاعات حاصل از ممیزی انرژی در ساختمان‌های اداری به کمک تئوری راف تحلیل شده است. تئوری راف به‌عنوان نوعی ابزار

- 
1. Energy Cultures framework
  2. Cognitive norms
  3. Material culture
  4. Energy practices

محاسباتی جدید برای برخورد با شرایط مبهم و عدم قطعیت معرفی شده است. این تئوری را می‌توان برای تحلیل اطلاعات نادقیق، متناقض و ناکامل به کار گرفت. با توجه به اینکه هدف اصلی این مقاله ارائه مدل قوانین به کمک تئوری مجموعه‌های راف است، در بخش بعدی تئوری مجموعه‌های راف و مفاهیم اساسی مرتبط با آن تشریح شده و پس از آن پیشینه تحقیق در خصوص این تئوری و به کارگیری آن در زمینه انرژی مرور خواهد شد. در ادامه، فرایند تولید قوانین به کمک RST و چگونگی به کارگیری آن در این تحقیق توضیح داده شده است. در نهایت خروجی‌های اجرای مدل به کمک نرم‌افزار ROSETTA و نتیجه‌گیری ارائه شده است.

### پیشینه نظری پژوهش

#### تئوری مجموعه‌های راف

RST<sup>۱</sup> به عنوان ابزاری قدرتمند برای استنتاج داده‌ها و تحلیل و پیش‌بینی تصمیمات، داده‌کاوی، ساخت سیستم‌های خبره و سیستم‌های پشتیبان تصمیم‌گیری و بسیاری از زمینه‌های دیگر کاربرد دارد. نقطه آغاز RST مجموعه داده‌هاست که به طور معمول در قالب یک جدول سازمان داده شده و تحت نام سیستم اطلاعاتی یا پایگاه داده معرفی می‌شوند. به بیان دیگر  $S = (U, A, V, f)$  یک سیستم اطلاعاتی است که  $U$  مجموعه مرجع و  $A$  مجموعه مشخصه‌های آن است.

$$U = \{U_1, U_2, \dots, U_{|U|}\}$$

$$A = \{a_1, a_2, \dots, a_{|A|}\}$$

$V = \cup V_a$  و  $V_a$  دامنه مشخصه  $a$  است و  $f: U \times A \rightarrow V$  تابعی است که برای  $\forall a \in A, \forall x \in U$  ارزش مشخصه هر یک از اجزا را به صورت  $f(x, a) \in V_a$  تخصیص می‌دهد (جیان، لیو و لین، ۲۰۱۱).

در عمل به طور معمول براساس نتایج شناخته شده اطلاعات را دسته‌بندی می‌کنند که این دسته‌بندی را با یک شاخص به نام مشخصه تصمیم<sup>۲</sup> نشان می‌دهیم. سیستم اطلاعاتی دارای این شاخص را سیستم تصمیم‌گیری می‌نامند. اگر یک سیستم اطلاعاتی به شکل  $S = (U, A, V, f)$  داشته باشیم، آن‌گاه شاخص‌های موجود در مجموعه  $A$  را می‌توان به دو

1. Rough Set Theory  
2. Decision attribute

زیرمجموعه جدا از هم تقسیم کرد که به یک گروه مشخصه‌های موقعیتی<sup>۱</sup> یا C و به گروه دیگر مشخصه‌های تصمیم یا D گفته می‌شود و خواهیم داشت  $A = C \cup D$ ،  $C \cap D = \emptyset$ ؛ در این صورت S را جدول تصمیم می‌نامیم.

سیستم اطلاعاتی  $S = (U, A, V, f)$  را در نظر بگیرید که برای هر  $B \subseteq A$  یک رابطه هم‌ارزی به شکل زیر برقرار است:

$$IND_S(B) = \{(x, x') \in U^2 : \forall a \in B, a(x) = a(x')\} \quad \text{رابطه (۱)}$$

اگر  $(x, x') \in IND_S(B)$  باشد، آن‌گاه با در نظر گرفتن شاخص‌های B اعضای  $x, x'$  از یکدیگر تشخیص‌ناپذیرند. به بیان دیگر این دو عضو را نمی‌توان از طریق شاخص‌های موجود در مجموعه B از هم متمایز کرد. کلاس‌های هم‌ارز مربوط به رابطه IND را به شکل  $[x]_B$  نشان می‌دهند.

اگر سیستم اطلاعاتی  $S = (U, A)$  را داشته باشیم و  $B \subseteq A$ ،  $X \subseteq U$  می‌توانیم X را تنها با اطلاعات موجود در B و با ساختن تقریب پایین و بالای B از X تقریب بزنیم که داریم:

$$\underline{B}X = \underline{apr}_B(X) = \{x : [x]_B \subseteq X\} \quad \text{رابطه (۲)}$$

$$\overline{B}X = \overline{apr}_B(X) = \{x : [x]_B \cap X \neq \emptyset\} \quad \text{رابطه (۳)}$$

اعضای  $\underline{B}X$  را می‌توان با قطعیت و بر پایه دانش موجود در B عضو X دانست، درحالی‌که اعضای  $\overline{B}X$  را می‌توان اعضای محتمل یا ممکن X دانست.

تعیین برخی ویژگی‌های مهم سیستم اطلاعاتی براساس رابطه «تمییزناپذیر»<sup>۲</sup> امکان‌پذیر است. اگر تعداد اعضای همسانی که از مجموعه شاخص‌های A استخراج می‌شود با تعداد اعضای همسان حاصل از  $A - a_i$  برابر باشد، آن‌گاه شاخص  $a_i$  را شاخص زائد یا تکراری<sup>۳</sup> می‌نامند و در غیر این صورت به آن شاخص ضروری<sup>۴</sup> گفته می‌شود.

اگر D و C دو زیرمجموعه از A باشند، آن‌گاه D را با شدت  $k$  ( $0 \leq k \leq 1$ ) وابسته به C می‌دانیم و به شکل  $C \Rightarrow_k D$  نشان می‌دهیم که مقدار k برابر است با:

1. Conditional attribute  
 2. Indiscernibility  
 3. Redundant  
 4. Indispensable

$$k = \gamma(C, D) = \frac{|POS_C(D)|}{|U|} \quad \text{رابطه ۴}$$

$$POS_C(D) = \bigcup_{X \in U/D} C(X) \quad \text{رابطه ۵}$$

اگر  $k = 1$  باشد  $D$  کاملاً به  $C$  وابسته است و در غیر این صورت وابستگی نسبی و با شدت  $k$  بین این دو مجموعه وجود دارد.

در یک سیستم اطلاعاتی به شکل  $S = (U, A, V, f)$ ,  $p \subseteq A$ ,  $X \subseteq U$  می‌توان دقت تخمین  $X$  را به شکل زیر تعریف کرد:

$$\alpha_p(X) = \frac{|apr_p(X)|}{|\overline{apr}_p(X)|} \quad \text{رابطه ۶}$$

دقت تخمین نشان‌دهنده میزان فهم حاصل از  $X$  و براساس دانش موجود است. روشن است که  $0 \leq \alpha_p(X) \leq 1$ . اگر  $\alpha_p(X) = 1$  باشد، آن‌گاه مجموعه  $X$  یک مجموعه کریسپ است و اگر  $\alpha_p(X) < 1$  باشد، آن‌گاه مجموعه  $X$  یک مجموعه راف است.

اگر یک سیستم اطلاعاتی به شکل  $S = (U, A)$  داشته باشیم، یک «بی‌زائده» برای مجموعه  $S$  عبارت است از مجموعه‌ای کمینه از شاخص‌ها به شکل  $B \subseteq A$  به‌قسمی که  $IND_S(B) = IND_S(A)$ . به بیان دیگر یک «بی‌زائده»، مجموعه‌ای کمینه از شاخص‌هاست که دسته‌بندی سیستم را حفظ می‌کند و توانایی دسته‌بندی‌ای نظیر کل شاخص‌ها را داراست. از آنجاکه ممکن است بیش از یک «بی‌زائده» در جدول تصمیم وجود داشته باشد، نقطه مشترک همه بی‌زائده‌ها را هسته جدول تصمیم می‌نامند که به شکل زیر نشان داده می‌شود:

$$CORE(P) = \bigcap_{R_i \in RED(P)} R_i \quad i = (1, 2, \dots, n) \quad \text{رابطه ۷}$$

در بسیاری از سیستم‌های یادگیری، استنتاج قوانین یکی از وظایف ضروری است. مجموعه تمامی مشخصه‌های موقعیتی را کلاس‌های موقعیتی  $S^1$  می‌نامند و به شکل  $X_i (i = 1, 2, \dots, n)$  نشان می‌دهند. مجموعه مشخصه‌های تصمیم را کلاس‌های تصمیم  $S^2$  می‌نامند و به شکل  $Y_j (j = 1, 2, \dots, n)$  نشان می‌دهند. قوانین تصمیم  $(C, D)$  را به شکل  $\{r_{ij}\}$  نشان می‌دهیم.

- 
1. Condition classes
  2. Decision classes

رابطه ۸)  $\forall i, j, \text{ if } X_i \subseteq Y_j$

آن گاه قانون  $r_{ij}$  قطعی است و در غیر این صورت غیر قطعی است.  
نمایش یک قانون به صورت زیر است:

رابطه ۹)  $IF, f(x, q_1) = r_{q1} \wedge f(x, q_2) = r_{q2} \wedge \dots \wedge f(x, q_p) =$

$r_{qp} THEN x \in Y_{j1} \vee Y_{j2} \vee \dots \vee Y_{jk} \{q_1, q_2, \dots, q_p\} \subseteq C;$

$(r_{q1}, r_{q2}, \dots, r_{qp}) \in V_{q1} \times V_{q2} \times \dots \times V_{qp}.$

در این تحقیق به منظور استخراج قوانین، از الگوریتم‌های مختلفی استفاده شده که در بخش‌های بعدی به تفصیل توضیح داده شده‌اند.

### پیشینه تحقیق

بیشتر مدل‌های سنتی ایجاد شده در زمینه مصرف انرژی در بخش خانگی، در پی تعیین محرک‌های رفتاری و رابطه بین این محرک‌ها هستند. هیچکاک برای تشریح زیرسیستم انسانی و زیرسیستم فیزیکی یک خانوار، از چارچوبی بر پایه سیستم استفاده کرد که از دیدگاه علم جامعه‌شناسی و علم مهندسی به مطالعه زیرسیستم‌ها می‌پردازد. اساس مدل او آن است که مصرف انرژی از رفتار ناشی می‌شود و در کنار آن ویژگی‌های فیزیکی خانه نیز بر مصرف اثرگذار است. علوم اجتماعی به مقوله مصرف انرژی به عنوان یک رویداد انسان محور نگاه می‌کنند و بر زیرسیستم انسانی و سیستم‌های اقتصادی و فرهنگی در محیط تمرکز دارند. مدل‌های فرهنگی تلاش دارند رابطه بین جنبه‌های اساسی رفتارهای اجتماعی و مصرف انرژی را تشریح کنند (هیچکاک، ۱۹۹۳). در این قسمت برخی تحقیقات در زمینه بررسی مصرف انرژی از نظر آسایش و ساکنان و همچنین تحقیقات در زمینه به کارگیری RST بررسی شده است.

در تحقیقی رضوی، مهرگان، شکوری و کریمی (۱۳۹۱) به خوشه‌بندی ساختمان‌های اداری پرداختند و داده‌های حاصل از ممیزی انرژی ساختمان‌ها را به کمک تئوری مجموعه‌های خاکستری تحلیل کردند. براساس نتایج به کارگیری روش خوشه‌بندی با وزن‌های ثابت و متغیر، از بین ۳۸ ساختمان مورد بررسی، تنها ۳ ساختمان تقریباً استانداردند و ۲۴ ساختمان انحراف متوسط دارند. به بیان دیگر کمتر از ۱۰ درصد ساختمان‌های مورد بررسی تابع وزارت نفت با استانداردهای مبحث ۱۹ مطابقت نسبی دارند و حدود ۳۰ درصد از این ساختمان‌ها کاملاً غیراستانداردند.

هدف مقاله احمدوند، قادری و بوستان‌پور (۲۰۱۰) تحلیل مصرف انرژی الکتریکی در بخش خانگی ایران با استفاده از رویکرد پویایی سیستم است که به این منظور مصرف انرژی با به کارگیری مفاهیم SD شبیه‌سازی شده است. از این رو ابتدا عوامل اثرگذار بر مصرف و چرخه‌های علی - معلولی با توجه به ادبیات موضوع مشخص شده است. سؤال اصلی تحقیق این است که سیاست‌های مدیریت مصرف، چگونه و تا چه اندازه بر مصرف اثرگذار است و آیا می‌توان اثر این تصمیمات را با روابط خطی پیش‌بینی کرد.

میشالیک و میلزارسکی (۱۹۹۶) روش جدیدی را برای ارزیابی مصرف انرژی در ساختمان‌های اداری با استفاده از اطلاعات حاصل از صورتحساب‌های پرداختی معرفی کرده و نتایج را با نتایج مدل PRISM مقایسه کرده‌اند.

یو و همکاران (۲۰۱۱) برای نشان دادن تأثیر رفتار ساکنان بر مصرف انرژی ساختمان، روش جدیدی ارائه کردند که برپایه فنون داده‌کاوی بنا شده است. برای نشان دادن ارتباط بین دو فاکتور و محاسبه ضریب وزنی هر شاخص در خوشه‌بندی، از درجه همبستگی خاکستری<sup>۱</sup> استفاده شده است. نتایج به دست آمده، در اولویت‌بندی فعالیت‌هایی که سبب تغییر رفتار ساکنان در جهت کاهش مصرف انرژی می‌شود استفاده شده و در مدل‌سازی رفتار ساکنان به کار رفته است.

RST در زمینه‌های مختلفی نظیر تشخیص پزشکی، قابلیت اطمینان در مهندسی، سیستم‌های خبره، ارزیابی ریسک و رشکستگی، پیش‌بینی شکست یا موفقیت کسب و کار، تحلیل تقاضای حمل و نقل، قیمت سهام، بازار بیمه و پیشگیری از تصادفات استفاده شده است. در تحقیق لیو و لین (۲۰۱۰) از DRSA<sup>۲</sup> برای ایجاد قوانینی به منظور تعیین نگرش مشتریان و وفاداری آنها استفاده شده تا از این طریق استراتژی‌هایی برای دستیابی به مشتریان جدید و نگهداشت مشتریان پرارزش توسعه یابد. در واقع شرکت‌ها به جای هدفگذاری‌های یکسان بر روی همه مشتریان یا ارائه پیشنهادهای انگیزاننده یکسان برای همه آنها می‌توانند برپایه نیازهای شخصی یا رفتارهای خرید مشتریان هدفگذاری کنند.

در این تحقیق از هر پاسخگو خواسته شد تا سطح رضایتمندی از هر شاخص و تصمیمات خرید از خط هوایی خاص را مشخص کند. ۴۵۱ پاسخگو این پرسشنامه را تکمیل کردند که نتیجه تحلیل شامل کیفیت تقریب، ایجاد قوانین، اعتبارسنجی قوانین و معناداری مشخصه‌های موقعیتی در مقاله آمده است. برای امکان‌سنجی قوانین تصمیم‌تولیدی از تکنیک اعتبارسنجی موازی استفاده شده است. در این تحقیق از «بی‌زائده» و هسته استفاده نشد و همه مشخصه‌های

1. Grey relational grades  
2. Dominance-based Rough Set Approach

انتخابی به عنوان فاکتورهای مهم برای افزایش کیفیت تقریب به کار گرفته شد، چراکه تنها شاخص‌های بااهمیت را شامل می‌شدند. نتیجه این مطالعه، تعیین دو شاخص ایمنی و قیمت به عنوان شاخص‌های ارجح در تصمیم مشتریان بوده است.

زو و همکاران به انتخاب و ارزیابی توزیع‌کنندگان با اطلاعات کیفی و چگونگی تحلیل این اطلاعات با استفاده از ابزاری غیر از تکنیک‌های استاندارد آماری پرداختند. در این مقاله متدی بر پایه RST معرفی شده که به عنوان ابزاری قدرتمند برای انتخاب توزیع‌کننده ارجح به کار گرفته شده است. در رویکرد پیشنهادی، وزن هر یک از ورودی‌ها نیز وارد مدل شده تا کیفیت قوانین استخراجی افزایش یابد (زو و همکاران، ۲۰۱۱).

متد RST در تحقیق گالوجان و یان برای ارزیابی اعتبار پیمانکاران به کار رفته است. این متد از پایگاه داده‌های پیمانکاران و از الگوریتم ژنتیک موجود در نرم‌افزار ROSETTA استفاده کرده است. ۱۰۸ مشخصه از وضعیت سازمان، کیفیت، جوایز و مسائل ثبت شده در پایگاه داده انتخاب شده و بسیاری از آنها پس از نظرخواهی از خبرگان حذف شدند. از تکنیک‌های پیش‌پردازش نظیر تحلیل روابط برای حذف مشخصه‌های مرتبط به یکدیگر استفاده و در نهایت ۱۶ مشخصه موقعیتی به عنوان متغیر مستقل انتخاب شد. مشخصه تصمیم در این تحقیق، «وضعیت اعتبار» انتخاب شده و مقدار آن از عملکرد گذشته یا ارزیابی‌های سنتی به دست آمده است (گالوجان و یان، ۲۰۰۶).

با توجه به کاربردهای RST در پیشینه تحقیق، می‌توان از ابزارهای این تئوری در تحلیل داده‌های حاصل از ممیزی انرژی ساختمان‌های اداری استفاده کرد و علاوه بر آن بین شاخص‌های فنی و جنبه‌های غیرفنی اثرگذار بر مصرف انرژی ساختمان‌ها ارتباط برقرار کرد که در بخش بعد این کار صورت گرفته است.

### روش‌شناسی تحقیق

تحقیق حاضر از نظر هدف، توسعه‌ای-کاربردی است. از سوی دیگر این پژوهش، تحقیقی کاربردی است، چراکه از الگوی پیشنهادی به صورت اجرایی و عملی در حوزه مصارف انرژی سازمان‌ها می‌توان استفاده کرد. از طرفی، از آنجاکه هدف این تحقیق، توصیف شرایط و پدیده‌های مورد بررسی، به منظور شناخت بیشتر شرایط موجود و کمک به فرایند تصمیم‌گیری است، این تحقیق را براساس چگونگی به دست آوردن داده‌های مورد نظر یا ماهیت تحقیق می‌توان توصیفی به‌شمار آورد. قلمرو مکانی این تحقیق سازمان‌های تابع وزارت نفت واقع در شهر تهران هستند که برای آنها ممیزی انرژی صورت گرفته و گزارش‌های این ممیزی در اختیار



محقق قرار گرفته است. از نظر قلمرو زمانی، ساختمان‌هایی که ممیزی انرژی آنها از سال ۱۳۸۵ توسط شرکت‌های مشاور شروع شده و چارچوب به نسبت یکسانی دارند مدنظر قرار گرفته‌اند. به‌طور خلاصه گام‌های اساسی برای استخراج قوانین تصمیم در تئوری راف به شرح زیر است:

۱. ساختن سیستم تصمیم براساس مشخصه‌های موقعیتی و مشخصه‌های تصمیم؛
۲. استخراج تقریب بالا، پایین، بی‌زائده‌ها و هسته از جدول تصمیم به کمک الگوریتم‌های موجود؛
۳. تولید مجموعه قوانین تصمیم براساس استراتژی‌های مختلف؛
۴. اعتبارسنجی قوانین به دست آمده.

در این تئوری به مجموعه قوانین حاصل، مدل قوانین گفته می‌شود که شرحی کلی از یک سیستم اطلاعاتی ارائه می‌کند. این مدل دو هدف زیر را دنبال می‌کند:

- اهداف پیش‌بینی: از این مدل می‌توان برای پیش‌بینی مشاهدات جدید استفاده کرد؛
- اهداف توصیفی: این مدل کمک می‌کند تا بفهمیم چه چیزی موجب تمایز مشاهدات می‌شود.

ساخت مدل قوانین در این تحقیق با در نظر گرفتن هر دو هدف بالا صورت گرفته است. در ادامه هر یک از گام‌های فوق با در نظر گرفتن هدف این تحقیق، تشریح شده است.

### ساخت سیستم تصمیم

در این تحقیق به منظور جمع‌آوری داده‌های مورد نیاز، پس از بررسی منابع کتابخانه‌ای و با مراجعه به مراکز پژوهشی وزارت نفت نظیر مؤسسه مطالعات انرژی و سازمان بهینه‌سازی مصرف سوخت، گزارش‌های ممیزی انرژی تعدادی از ساختمان‌ها که در حیطه موضوعی این تحقیق قرار داشتند بررسی شد. پس از مطالعه دقیق این گزارش‌ها، مشخصه‌های موقعیتی (C) و مشخصه تصمیم (D) مورد نیاز برای این تحقیق با توجه به داده‌های موجود در همه آنها شناسایی شد. نام هر مشخصه، واحد اندازه‌گیری، نوع مشخصه، دامنه مقادیر ممکن و تعداد مشاهدات بدون مقدار برای هر مشخصه در جدول ۱ نشان داده شده است. مشاهدات در این تحقیق (U) ساختمان‌های اداری ممیزی شده (۴۱ ساختمان) هستند. داده‌های مربوط به مشخصه‌ها (Va) برای هر یک از مشاهدات از گزارش‌های ممیزی انرژی ساختمان به دست آمده است.

در جدول ۱ منظور از فضای کنترل نشده بخش‌هایی از فضای ساختمان مانند پارکینگ‌ها، راه‌پله‌ها و نظایر آن است که مورد سرمایه‌گذاری و گرمایش قرار نمی‌گیرند. ضریب بار هر یک از جداره‌های ساختمان از حاصل ضرب ضریب انتقال حرارت آن جزء در مساحتش به دست می‌آید. ضریب انتقال حرارت سطحی، برابر است با توان حرارتی منتقل شده از آن سطح، با مساحت ۱

متر مربع در صورتی که اختلاف دمای داخل و خارج ۱ درجه باشد. بنابراین ضریب بار یک سطح، کل حرارت انتقالی از آن سطح را زمانی که اختلاف دمای داخل و خارج ۱ درجه باشد نشان می‌دهد.

جدول ۱. مشخصه‌های سیستم تصمیم

مشخصه	عنوان	واحد اندازه‌گیری	نوع مشخصه	دامنه مقادیر	مشاهدات بدون مقدار
C <sub>1</sub>	تعداد طبقات	تعداد	عدد صحیح <sup>۱</sup>	[۱، ۲۱]	۲
C <sub>2</sub>	متراژ ساختمان	متر مربع	اعشار <sup>۲</sup>	[۳۲۵، ۲۳۸۰۰]	۱
C <sub>3</sub>	عمر تقریبی ساختمان	سال	عدد صحیح	[۲، ۵۴]	۱۴
C <sub>4</sub>	متراژ فضای کنترل‌نشده ساختمان	متر مربع	اعشار	[۷۰/۱۴، ۸۰/۲۳۴۸]	۹
C <sub>5</sub>	تعداد ساکنان	نفر	عدد صحیح	[۴، ۱۳۰۰]	۳
C <sub>6</sub>	ضریب بار دیوارها	وات بر کلوبین	اعشار	[.۴۷، ۳۲/۴]	۱
C <sub>7</sub>	ضریب بار پنجره‌ها	وات بر کلوبین	اعشار	[۳، ۸/۷]	۱
C <sub>8</sub>	ضریب بار درها	وات بر کلوبین	اعشار	[۲، ۹۷/۵]	۱
C <sub>9</sub>	ضریب بار سقف	وات بر کلوبین	اعشار	[۵/۰، ۲]	۱
C <sub>10</sub>	ضریب بار کف	وات بر کلوبین	اعشار	[.۲، ۴۲/۲]	۲۴
C <sub>11</sub>	نوع پنجره	-	اسمی	تک‌جداره- فلزی دوجداره- فلزی تک‌جداره- آلومینیومی دوجداره- آلومینیومی تک‌/دوجداره- آلومینیومی تک‌جداره- فلزی / آلومینیومی	۱
D	سطح آسایش کارکنان	-	اسمی	مناسب / متوسط / نامناسب	۶

در دنیای واقعی به‌طور معمول مقدار برخی داده‌ها مشخص نیست. این داده‌ها ممکن است بر نتایج حاصل از تکنیک‌های تئوری مجموعه‌های راف اثرگذار باشند. به همین دلیل قبل از شروع فرایند تولید قوانین باید روش برخورد با مقادیر نامعلوم مشخص شود. روش‌های متنوعی برای

1. Integer  
2. Float

تکمیل داده‌ها وجود دارد که مرسوم‌ترین آنها عبارتند از حذف داده‌های ناقص<sup>۱</sup>، میانگین/مد<sup>۲</sup>، میانگین/مد شرطی<sup>۳</sup>، تکمیل ترکیبی<sup>۴</sup> و تکمیل ترکیبی شرطی<sup>۵</sup>. در این تحقیق به دلیل تعداد کم داده‌ها، امکان به کارگیری روش حذف داده‌های ناقص وجود ندارد؛ همچنین به دلیل تنوع دامنه مقادیر هر یک از مشخصه‌ها، به کارگیری روش تکمیل ترکیبی به بزرگ شدن بیش از حد سیستم اطلاعاتی منجر می‌شود. از این رو از دو روش میانگین/مد شرطی و تکمیل ترکیبی شرطی استفاده شده است که پس از تکمیل داده‌ها به روش ترکیبی شرطی، ۴۱ مشاهده اصلی به ۱۹۰۸ مشاهده تبدیل شد (ن.ک: کریمی و صادقی مقدم، ۲۰۱۴).

در RST اگر مقادیر تخصیص داده شده به مشخصه‌ها پیوسته باشند، دامنه مقادیر ممکن، بسیار گسترده شود و در تولید قوانین با مشکل روبه‌رو خواهیم شد؛ چراکه بسیاری از اعضا در یک مشخصه با هم تفاوت پیدا می‌کنند و با اینکه ممکن است تفاوت دو عضو در آن مشخصه بسیار ناچیز باشد، ماهیت تمایز و همسانی در RST، این دو عضو را از هم تفکیک می‌کند. بنابراین قبل از تولید قوانین باید به کمک روش‌های مناسب، داده‌های عددی به بازه‌های گسسته<sup>۶</sup> تبدیل شوند.

مسئله تولید بازه‌های گسسته یک مسئله NP است و راه حل بهینه برای آن وجود ندارد؛ اما برخی از روش‌های متداول آن به شرح زیر است:

- Expert Discrete Method
- Equal Width Interval Method and Equal Frequency Interval Method
- The Most Subdivision Entropy Method
- Chimerge Method

برای اطلاع از الگوریتم هر یک از روش‌های فوق ن.ک: لودل و ویدمر (۲۰۰۰)؛ روییز، آنگولو و اگل (۲۰۰۸). در این تحقیق از روش آنتروپی برای گسسته‌سازی داده‌های هر مشخصه استفاده شده است (ن.ک: کلارک و بارتون، ۲۰۰۰).

## تولید بی‌زائده

یکی از مهم‌ترین جنبه‌ها در تحلیل جدول اطلاعاتی استخراج و حذف مشخصه‌های تکراری و همچنین شناسایی مهم‌ترین مشخصه از درون داده‌هاست. مشخصه‌های تکراری به آنهایی گفته

- 
1. Remove incompletes
  2. Mean/ Mode fill
  3. Conditional Mean/ Mode fill
  4. Combinatorial completion
  5. Conditional Combinatorial completion
  6. Discrete interval

می‌شود که حذف آنها هیچ تأثیری بر شدت وابستگی بین مشخصه‌های موجود و مشخصه تصمیم ندارد. شدت وابستگی اندازه‌ای است که قدرت تمییز بین مشاهدات را نشان می‌دهد. حداقل زیرمجموعه مشخصه‌ها را که شدت وابستگی آنها حداکثر است بی‌زائده می‌نامند. محاسبه هسته و بی‌زائده‌های یک جدول اطلاعاتی در واقع انتخاب مشخصه‌های مهم و مرتبط آن سیستم اطلاعاتی است. هر بی‌زائده در مجموعه بی‌زائده‌های تولیدشده دارای یک معیار پشتیبان<sup>۱</sup> است. این معیار نشان‌دهنده قدرت آن بی‌زائده است و ممکن است براساس الگوریتم مورد استفاده متفاوت تفسیر شود. مقدار این معیار برابر با نسبت درستنمایی آن بی‌زائده ضرب در ۱۰۰ است. الگوریتم‌های رایج برای محاسبه بی‌زائده‌ها شامل الگوریتم ژنتیک، الگوریتم جانسون<sup>۲</sup>، تک قانون هولت<sup>۳</sup>، بی‌زائده دستی<sup>۴</sup> و نظایر آن است (اسمیت و بال، ۲۰۰۳ و استارزیک، نلسون و استورتز، ۲۰۰۰). در این تحقیق از الگوریتم‌های جانسون و ژنتیک برای استخراج بی‌زائده‌های سیستم تصمیم استفاده شد. خروجی الگوریتم جانسون تنها یک مجموعه بی‌زائده است که پس از اجرای آن بر روی دو سیستم تصمیم تکمیل‌شده دو مجموعه زیر به‌عنوان بی‌زائده به‌دست آمد:

$$\{\text{«مترای فضای کنترل نشده (C4)»}, \text{«ضریب بار کف (C10)»}\} = \text{جدول میانگین/مد شرطی}$$

$$\{\text{ضریب بار کف (C10)}\} = \text{جدول ترکیبی شرطی}$$

بی‌زائده‌های حاصل از الگوریتم ژنتیک برای دو سیستم تصمیم تکمیل‌شده در جدول‌های ۲ و ۳ نشان داده شده است. مقدار  $\gamma(C, D)$  برای همه بی‌زائده‌های تولیدشده در هر دو روش برابر یک است.

جدول ۲. بی‌زائده‌های الگوریتم ژنتیک جدول میانگین/مد شرطی

تعداد مشخصه	مجموعه بی‌زائده	تعداد مشخصه	مجموعه بی‌زائده
۳	{C5, C6, C10}	۲	{C4, C10}
۴	{C2, C6, C9, C11}	۳	{C2, C10, C11}
۴	{C2, C3, C9, C11}	۳	{C2, C9, C10}
۴	{C4, C6, C9, C11}	۳	{C5, C10, C11}
		۳	{C2, C6, C10}

1. Support count
2. Johnson's algorithm
3. Holte's 1R
4. Manual reducer

جدول ۳. بی‌زائده‌های الگوریتم زنتیک جدول ترکیبی شرطی

تعداد مشخصه	مجموعه بی‌زائده	تعداد مشخصه	مجموعه بی‌زائده
۳	{C۳, C۸, C۱۱}	۱	{C۴}
۳	{C۵, C۶, C۷}	۲	{C۷, C۱۰}
۳	{C۱, C۹, C۱۰}	۲	{C۲, C۶}
۳	{C۲, C۵, C۹}	۲	{C۲, C۱۰}
۳	{C۲, C۷, C۱۱}	۲	{C۵, C۱۰}
۳	{C۳, C۶, C۸}	۲	{C۳, C۱۰}
۳	{C۲, C۳, C۱۱}	۳	{C۱, C۸, C۱۰}
۳	{C۳, C۵, C۶}	۳	{C۶, C۷, C۸}
۴	{C۱, C۵, C۸, C۹}	۳	{C۵, C۶, C۹}
۴	{C۱, C۶, C۷, C۱۱}	۳	{C۵, C۷, C۱۱}
۴	{C۲, C۷, C۸, C۹}	۳	{C۵, C۷, C۹}
۴	{C۱, C۶, C۹, C۱۱}	۳	{C۱, C۶, C۱۰}
۴	{C۳, C۷, C۹, C۱۱}	۳	{C۳, C۸, C۹}
۴	{C۶, C۷, C۹, C۱۱}	۳	{C۲, C۳, C۹}

### استخراج مجموعه قوانین تصمیم بر اساس استراتژی‌های مختلف

برای استخراج قوانین تصمیم از مشخصه‌های درون هر بی‌زائده استفاده می‌شود. در سمت چپ قوانین تصمیم، ترکیبی از ارزش‌های مشخصه‌هاست که به‌طور معمول مشخصه تصمیم همه مشاهداتی که با این ترکیب همخوانی دارند برابر سمت راست قانون است. قوانین IF-THEN با خواندن ارزش هر یک از مشخصه‌های موجود در بی‌زائده و ارتباط دادن آن به یک یا چند گروه تصمیم ساخته می‌شوند. برای ارزیابی قوانین تولیدی از شاخص عمومیت قانون<sup>۱</sup> استفاده می‌شود که برابر درصد مشاهداتی است که ضمن تعلق به گروه تصمیم «بخش آن‌گاه» شرایط «بخش اگر» را نیز پوشش می‌دهند. همچنین استفاده از شاخص دقت قانون<sup>۲</sup> که برابر درصد مشاهداتی است که شرایط «بخش اگر» را دارند و به گروه تصمیم «بخش آن‌گاه» تعلق دارند نیز به ارزیابی قوانین کمک می‌کند (برای اطلاع از جزئیات الگوریتم استنتاج قوانین ن.ک: گریزمالا-باسه و استفانوفسکی، ۲۰۰۱؛ کریمی و صادقی‌مقدم، ۲۰۱۴).

برای هر یک از جدول‌های تکمیل‌شده و هر یک از روش‌های تولید بی‌زائده، مجموعه قوانین به‌دست آمد که قسمتی از نتایج آن در جدول‌های ۴ تا ۷ نشان داده شده است.

1. Coverage  
2. Accuracy

## جدول ۴. مجموعه قوانین میانگین / جانسون

عمومیت	دقت	پنج قانون اول
۰/۵	۱	C۴ ([۳۶۲/۵۹, ۳۷۰/۹۶]) AND C۱۰ ([۱/۵۷, ۱/۷۱]) => D (good)
۰/۴	۱	C۴ ([۴۴۳۴/۵۰, *]) AND C۱۰ ([*, ۱/۱۰]) => D (bad)
۰/۲۳	۱	C۴ ([*, ۳۶۲/۵۹]) AND C۱۰ ([*, ۱/۱۰]) => D (so so)
۰/۲۳	۱	C۴ ([*, ۳۶۲/۵۹]) AND C۱۰ ([۱/۲۸, ۱/۵۷]) => D (so so)
۰/۲۳	۱	C۴ ([۵۱۷/۲۵, ۷۵۰/۱۵]) AND C۱۰ ([۱/۲۸, ۱/۵۷]) => D (so so)
۱۸ قانون		کل قوانین

در جدول بالا منظور از قانون اول این است که اگر مقدار «مترائز فضای کنترل نشده ساختمان» بین ۳۶۲/۵۹ و ۳۷۰/۹۶، و مقدار مشخصه «ضریب بار کف» بین ۱/۵۷ و ۱/۷۱ باشد، وضعیت آسایش کارکنان «مناسب» است. در قانون دوم مشاهده می شود که اگر مقدار «مترائز فضای کنترل نشده ساختمان» بیش از ۴۴۳۴/۵۰، و مقدار مشخصه «ضریب بار کف» تا ۱/۱۰ باشد وضعیت آسایش کارکنان «نامناسب» خواهد بود. در ضمن برای مقایسه این دو قانون می توان از شاخص عمومیت استفاده کرد که عمومیت قانون اول بیش از قانون دوم است.

## جدول ۵. مجموعه قوانین میانگین / GA

عمومیت	دقت	پنج قانون اول
۰/۵۰	۱	C۴ ([۳۶۲/۵۹, ۳۷۰/۹۶]) AND C۱۰ ([۱/۵۷, ۱/۷۱]) => D(good)
۰/۴۰	۱	C۲ ([۱۷۹۵/۸۰, ۲۱۳۴۶/۵۰]) AND C۹ ([*, ۱/۴۴]) AND C۱۰ ([*, ۱/۱۰]) => D(bad)
۰/۴۰	۱	C۴ ([۴۴۳۴/۵۰, *]) AND C۱۰ ([*, ۱/۱۰]) => D(bad)
۰/۴۰	۱	C۲ ([۱۷۹۵/۸۰, ۲۱۳۴۶/۵۰]) AND C۶ ([*, ۱/۸۶]) AND C۱۰ ([*, ۱/۱۰]) => D(bad)
۰/۳۱	۱	C۵ ([*, ۲۴۳]) AND C۱۰ ([۱/۲۸, ۱/۵۷]) AND C۱۱(۱- iron) => D(so so)
۲۸۲ قانون		کل قوانین

## جدول ۶. مجموعه قوانین ترکیبی شرطی / جانسون

عمومیت	دقت	پنج قانون اول
۰/۷۵	۱/۰۰	C۴ ([۱۰۹/۱۵, ۱۳۶/۰۰]) => D (bad)
۰/۵۴	۱/۰۰	C۴ ([۲۲۴/۰۰, ۳۹۱/۲۵]) => D (so so)
۰/۴۰	۱/۰۰	C۴ ([۴۱۵/۱۵, ۵۱۷/۲۵]) => D (good)
۰/۲۵	۱/۰۰	C۴ ([۱۳۶/۰۰, ۱۷۳/۰۰]) => D (so so)
۰/۲	۱	C۴ ([*, ۲۶/۳۵]) => D(good)
۱۴ قانون		کل قوانین

جدول ۷. مجموعه قوانین ترکیبی شرطی / GA

پنج قانون اول	دقت	عمومیت
$C6 ([1/41, 1/86]) \text{ AND } C7 ([6/34, *]) \text{ AND } C8 ([*, 4/55]) \Rightarrow D (\text{good})$	۱	۰/۷۶
$C5 ([49, 115]) \text{ AND } C7 ([4/24, 4/28]) \text{ AND } C11 (1 - \text{iron}) \Rightarrow D (\text{bad})$	۱	۰/۷۵
$C5 ([49, 115]) \text{ AND } C6 ([2/55, 3/45]) \text{ AND } C9 ([1/48, 1/75]) \Rightarrow D (\text{bad})$	۱	۰/۷۵
$C6 ([2/55, 3/45]) \text{ AND } C7 ([4/24, 4/28]) \text{ AND } C8 ([5/57, 5/65]) \Rightarrow D (\text{bad})$	۱	۰/۷۵
$C2 ([1434/00, 1658/50]) \text{ AND } C6 ([2/55, 3/45]) \Rightarrow D (\text{bad})$	۱	۰/۷۵
کل قوانین		۱۲۶۳ قانون

همان‌گونه که مشاهده می‌شود بیشترین تعداد قانون مربوط به جدول تکمیل‌شده ترکیبی شرطی است که به کمک ۲۸ بی‌زائده حاصل از الگوریتم ژنتیک، ۱۲۶۳ قانون تولید کرده است. همچنین کمترین تعداد قانون مربوط به جدول تکمیل‌شده ترکیبی شرطی است که با توجه به یک بی‌زائده حاصل از الگوریتم جانسون، ۱۴ قانون تولید کرده است. به‌منظور مقایسه چهار مدل به‌دست‌آمده، باید اعتبار هر یک از مدل‌ها در پیش‌بینی مشخصه تصمیم محاسبه شود.

### اعتبارسنجی قوانین به‌دست‌آمده

برای اعتبارسنجی قوانین به‌دست‌آمده باید یک جدول تصمیم با مجموعه مرجع  $U$  را به دو جدول تصمیم تصادفی و غیرمرتبط به هم با مجموعه‌های مرجع  $U_1, U_2$  تقسیم کرد، به طوری که:

$$U = U_1 \cup U_2, \quad U_1 \cap U_2 = \emptyset \quad (\text{رابطه } 10)$$

به بیان دیگر برای اعتبارسنجی باید داده‌ها به دو گروه یادگیری و کنترل تقسیم شوند تا به کمک داده‌های موجود در گروه یادگیری، قوانین استخراج شده و به کمک داده‌های موجود در گروه کنترل، نتایج حاصل از قوانین آزمون شوند.

در عمل باید داده‌ها را به‌طور تصادفی به گروه یادگیری و کنترل تقسیم کرد. اگرچه با تقسیم داده‌ها به دو گروه می‌توان قدرت پیش‌بینی و اعتبار مدل را سنجید، تقسیم‌بندی متفاوت داده‌ها ممکن است به نتایج متفاوت در شاخص‌هایی مانند «دقت مدل» بینجامد؛ بنابراین باید مدل برای گروه‌های یادگیری و کنترل متفاوت اجرا شود تا واریانس شاخص‌های عملکردی متفاوت آن استخراج گردد. از آنجاکه به‌طور معمول در عمل مشاهدات محدودی در دسترس است،

اعتبارسنجی موازی<sup>۱</sup> بهترین روش اعتبارسنجی است. اعتبارسنجی k تایی<sup>۲</sup> روشی است که نمونه‌ها را به k زیرمجموعه هم‌اندازه تقسیم می‌کند و یک زیرمجموعه را گروه کنترل و دیگر زیرمجموعه‌ها را گروه یادگیری در نظر می‌گیرد. این کار تکرار می‌شود تا هر زیرمجموعه یک بار به‌عنوان گروه کنترل و k-1 بار به‌عنوان گروه یادگیری انتخاب شود. برای اعتبارسنجی موازی در نرم‌افزار ROSETTA باید الگوریتم‌هایی که در هر تکرار استفاده می‌شود، به‌صورت اسکرپت‌های متنی نوشته شده و به‌عنوان ورودی وارد نرم‌افزار شود تا براساس آن اعتبارسنجی صورت پذیرد. در این تحقیق، پنج تکرار برای اعتبارسنجی موازی در نظر گرفته شده است. نتایج این گروه‌بندی به‌عنوان معیاری برای بررسی کیفیت قوانین حاصل و اعتبارسنجی مدل به‌کار می‌رود (کریمی و صادقی مقدم، ۲۰۱۴؛ ژونگ، دونگ و اوهسونکا، ۲۰۰۱).

فرایند اعتبارسنجی برای قوانین حاصل از هر دو جدول تکمیل شد و هر دو روش بی‌زائده انجام گرفت. در هر تکرار فرایند اعتبارسنجی یک جدول متقابل ارائه می‌شود. سطرهای این جدول معرف مقادیر واقعی مشخصه تصمیم و ستون‌ها معرف مقادیر پیش‌بینی شده توسط قوانین هستند. خانه آخر این ماتریس دقت قوانین را نشان می‌دهد. در اعتبارسنجی موازی، متوسط دقت در k تکرار به‌دست می‌آید. در جدول ۸ یک نمونه از ۲۰ جدول متقابل به‌دست‌آمده ارائه شده است. این جدول مربوط به تکرار اول اعتبارسنجی قوانین مربوط به جدول ترکیبی و بی‌زائده جانسون است. خلاصه نتایج اعتبارسنجی موازی در جدول ۹ ارائه شده است.

همان‌گونه که در جدول ۹ مشاهده می‌شود از بین چهار مدل تولیدشده، بیشترین دقت مربوط به جدول تکمیل ترکیبی شرطی و روش جانسون است. به بیان دیگر اگرچه تعداد قوانین تولیدشده در این مدل کمتر از سایر مدل‌هاست، نتایج اعتبارسنجی موازی نشان می‌دهد که دقت قوانین به‌دست‌آمده در این مدل از سایر مدل‌ها بیشتر است.

جدول ۸. نمونه جدول متقابل

مقادیر پیش‌بینی					
متوسط	مطلوب	نامطلوب	نامشخص	دقت	
۶	۰	۰	۰	۱۰۰	متوسط
۰	۰	۰	۱	۰	مطلوب
۰	۰	۱	۰	۱۰۰	نامطلوب
۰	۰	۰	۰	۰	نامشخص
۱۰۰	-	۱۰۰	۰	۵/۸۷	دقت

1. Cross validation
2. K-fold cross validation



جدول ۹. نتایج اعتبارسنجی موازی

جدول	بی‌زائده	حداقل دقت	حداکثر دقت	متوسط دقت
میانگین / مد شرطی	جانسون	۰/۳۷۵	۰/۶۲۵	۰/۴۶۴
میانگین / مد شرطی	ژنتیک	۰/۶۶۶	۰/۸۷۵	۰/۸۱
ترکیبی شرطی	جانسون	۰/۹۹۴	۱	۰/۹۹۸۹
ترکیبی شرطی	ژنتیک	۰/۹۹۲	۱	۰/۹۹۷۹

### نتیجه‌گیری

هدف از این تحقیق، تحلیل نتایج گزارش‌های ممیزی انرژی ساختمان‌های اداری وزارت نفت بود. در تحلیل این نتایج سعی شد به کمک ابزاری قدرتمند به نام مدل‌سازی راف قوانین حاکم بین مشخصه‌های فنی یک ساختمان و سطح آسایش کارکنان از منظر گرمایش و سرمایش استنتاج شود و درکی جامع از وضعیت مشخصه‌های مورد بررسی در این گزارش‌ها به دست آید. در فرایند مدل‌سازی راف از بین روش‌های مختلف تکمیل داده‌ها دو روش تکمیل براساس میانگین/مد شرطی و روش ترکیبی شرطی انتخاب شد. از سایر روش‌ها به دلیل برخی محدودیت‌ها استفاده نشد. از بین الگوریتم‌های متنوع گسسته کردن داده‌ها، الگوریتم آنتروپی انتخاب شد و از بین الگوریتم‌های تولید بی‌زائده دو روش قدرتمند و مرسوم جانسون و GA به کار گرفته شد و در مجموع چهار مدل قانون به دست آمد. بیشترین تعداد قوانین مربوط به مدل ترکیبی/GA با ۱۲۶۳ قانون و متوسط دقت ۹۹/۷ درصد و کمترین تعداد قوانین مربوط به مدل ترکیبی/جانسون با ۱۴ قانون و متوسط دقت ۹۹/۸ درصد است. ملاحظه می‌شود که در مدل دوم علی‌رغم تعداد کم قوانین، دقت قوانین و قدرت آنها برای پیش‌بینی مشاهدات جدید بیشتر است. هرچه تعداد عضوهای یک بی‌زائده کمتر باشد، آن بی‌زائده مطلوب‌تر است؛ از آنجاکه در روش جانسون تنها یک بی‌زائده تولید می‌شود، در مدل ترکیبی/جانسون این بی‌زائده تنها یک مشخصه دارد. به بیان دیگر با اطلاع از مشخصه C4 یعنی «مترائز فضای کنترل نشده» می‌توان مقدار مشخصه تصمیم D یعنی سطح آسایش کارکنان را با دقت ۹۹/۸ درصد پیش‌بینی کرد. بنابراین مهم‌ترین مشخصه سیستم تصمیم مورد بررسی C4 بوده و اگر در ممیزی انرژی ساختمان‌های اداری مقدار این مشخصه به دقت محاسبه شود نیازی به سنجش سطح آسایش کارکنان در قالب

پروژه‌های زمان‌بر و پرهزینه نیست و می‌توان به کمک مدل قوانین به دست آمده در این تحقیق مقدار آن را پیش‌بینی کرد.

اگرچه RST به عنوان یک ابزار محاسباتی جدید و قدرتمند برای برخورد با شرایط مبهم و عدم قطعیت معرفی شده است، اما به دلیل استفاده محدود در رشته‌هایی نظیر کامپیوتر و ریاضیات، در سایر علوم چندان به کار گرفته نشده است، از این رو این تحقیق می‌تواند منبعی مناسب برای شناخت اصول این تئوری، نرم‌افزار مربوط و چگونگی به کارگیری آنها باشد.

### **تشکر و قدردانی**

این مقاله از طرح پسادکترای به شماره ۹۲۰۱۹۰۶۶ استخراج شده که مورد حمایت مالی صندوق حمایت از پژوهشگران و فناوران کشور (INSF) قرار گرفته است.

### **References**

- Ahmadvand, A.M., Ghaderi, S.F. & Boustanpour, H.R. (2010). The analysis of residential energy consumption in Iran using system dynamic. *Second International Conference on Computer Modeling and Simulation*.
- Clarke, E.J. & Barton, B.A. (2000). Entropy and MDL Discretization of Continuous Variables for Bayesian Belief Networks. *International Journal of Intelligent Systems*, 15(1): 61-92.
- Gaojun, L. & Yan, Z. (2006). Credit Assessment of Contractors: A Rough Set Method, Tsinghua Science and Technology. *Tsinghua Science and Technology*, 11(4): 357-362.
- Grzymala-Busse, J.W. & Stefanowski, J. (2001). Three Discretization Methods for Rule Induction. *International Journal of Intelligent Systems*, 16 (1): 29-38.
- Hitchcock, G. (1993). An integrated framework for energy use and behaviour in the domestic sector, *Energy and Buildings*, 20(2):151-157.
- Jian, L., Liu, S. & Lin, Y. (2011). *Hybrid Rough Sets and Applications in Uncertain Decision-Making*. By Taylor and Francis Group, LLC. Auerbach Publications.

- Karimi, T. & Sadeghi Moghadam, M. (2014). *Rough Set and Grey Set Theory: principles, applications and software's*. First edition. Tehran: Mehraban nashr. (in Persian)
- Liu, S., Lin, Y., (2010). *Grey Systems Theory and Applications*. Springer-Verlag Berlin Heidelberg.
- Ludl, M.C. & Widmer, G. (2000). Relative Unsupervised Discretization for Association Rule Mining,” Proc. *Fourth European Conf. Principles of Data Mining and Knowledge Discovery (PKDD)*. 1910: 148-158.
- Lutzenhiser, L. (2008). Setting the stage: why behavior is important. Overview of address given to the behavior, *energy and climate change conference*, November 7–9 2007, Sacramento CA, delivered to California Senate June.
- Michalik, G. & Mielczarski, W. (1996). Modeling of Energy Use Patterns in the Residential Sector Using Linguistic Variables. *International Conference*, 28 Jan- 2 Feb. DOI: 10.1109/ISAP.1996.501082.
- Razavi, M., Mehregan, M., Shakori, H., Karimi, T., (2012) Analyzing Energy Consumption of Organizational Buildings Using Grey Set Theory. *Grey Systems: Theory and Application*, 4(3): 386 – 399.
- Ruiz, F.J., Angulo, C. & Agell, N., (2008), IDD: A Supervised Interval Distance-Based Method for Discretization. *Knowledge and Data Eng.*, 2 (9): 1230-1238.
- Smith, M. G., & Bull, L. (2003). Feature construction and selection using genetic programming and a genetic algorithm. *Proceedings of the 6th European Conference on Genetic Programming (EuroGP 2003)*, 229–237.
- Starzyk, J. A., Nelson, D. E., & Sturtz, K. (2000). A Mathematical Foundation for Improved Reduct Generation in Information Systems. *Journal of Knowledge and Information Systems*, 2 (2): 131-146.
- Stephenson, J., Barton, B., Carrington, G., Gnoth, D., Lawson, R., Thorsnes, P., (2010), Energy cultures: A framework for understanding energy behaviours. *Energy Policy*, doi:10.1016/j.enpol.2010.05.069.
- Yu, Z., Fung, C.M., Haghghat, F., Yoshino, H., Morofsky, E., (2011), A systematic procedure to study the influence of occupant behavior on building energy consumption, *Energy and Buildings*, 43, 1409–1417.

- Zhong, N., Dong, J., & Ohsuga, S., (2001). Using Rough Sets with Heuristics for Feature Selection. *Journal of Intelligent Information Systems*, 16 (3): 199–214.
- Zou, Z., Tseng, T.L., Sohn, H., Song, G., Gutierrez, R. (2011). A rough set based approach to distributor selection in supply chain management, *Expert Systems with Applications* 38 , 106–115.