



Rule Mining about the Relationship between Climatic Factors and the Number of Patients in a Hospital Using Classification Based on Multidimensional Association Rule Mining

Sima Hadadian

PhD Candidate, Department of Industrial Management, Faculty of Economic & Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: sima_hadadian@yahoo.com

Zahra Naji Azimi

*Corresponding Author, Associate Prof., Department of Industrial Management, Faculty of Economic & Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: znajiazimi@um.ac.ir

Naser Motahari Farimani

Assistant Prof., Department of Industrial Management, Faculty of Economic & Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: n.motahari@um.ac.ir

Behrooz Minaei Bidgoli

Associate Prof., Department of Artificial Intelligence, Faculty of Computer Engineering, Iran University of Science and Technology, Tehran, Iran. E-mail: b_minaei@iust.ac.ir

Abstract

Objective: There are many climatic factors affecting the number of patients in hospitals which generally tend to make a Non-optimal use of their facilities and human resources. This research is aimed at discovering hidden knowledge between climatic factors and the number of hospital patients using data mining techniques.

Methods: In this study, the relationship between climatic factors and the number of patients in Dr. Sheikh specialized pediatric hospital of Mashhad is investigated by classification based on multidimensional association rule mining. The number of patients in the nephrology, hematology, emergency and PICU department of this hospital have been considered separately, and consequently the relationship between the number of patients and the climatic factors such as air temperature, relative humidity, wind speed, air pressure and air pollution have been analyzed. This research has analyzed data gathered through a 19 month period and has been obtained by referring to the documents. In this research for feature selection, all subsets of climatic factors are searched and the effect of all subsets on the number of patients are evaluated using linear regression. Also for rule mining is used classification based on multidimensional association rule mining which is based on known Apriori algorithm.

Results: The results show different patterns that indicate the relationship between the number of patients in the hospital departments with the climatic factors.

Conclusion: This study is able to help analyze the relationship between the climatic factors and the number of patients in the hospital. Also, the rules will help managers make optimal planning for hospital resources according to the different number of patients.

Keywords: Classification based on multidimensional association rule mining, Apriori algorithm, Linear regression, Climatic factors, The number of hospital patients.

Citation: Hadadian, S., Naji Azimi, Z., Motahari Farimani, N., & Minaei Bidgoli, B. (2019). Rule Mining about the Relationship between Climatic Factors and the Number of Patients in a Hospital Using Classification Based on Multidimensional Association Rule Mining. *Industrial Management Journal*, 11(4), 575-599. (in Persian)

Industrial Management Journal, 2019, Vol. 11, No.4, pp. 575-599

DOI: 10.22059/imj.2019.283787.1007618

Received: May 18, 2019; Accepted: November 06, 2019

© Faculty of Management, University of Tehran



کاوش قواعد ارتباط‌دهنده عوامل جوئی با تعداد بیماران بیمارستان با استفاده از رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی

سیم‌ا حدادیان

دانشجوی دکتری، گروه مدیریت صنعتی، دانشکده علوم اداری و اقتصاد دانشگاه فردوسی مشهد، مشهد، ایران. رایانامه: sima_hadadian@yahoo.com

زهرا ناجی عظیمی

* نویسنده مسئول، دانشیار، گروه مدیریت صنعتی، دانشکده علوم اداری و اقتصاد دانشگاه فردوسی مشهد، مشهد، ایران. رایانامه: znajiazimi@um.ac.ir

ناصر مطهری فریمانی

استادیار، گروه مدیریت صنعتی، دانشکده علوم اداری و اقتصاد دانشگاه فردوسی مشهد، مشهد، ایران. رایانامه: n.motahari@um.ac.ir

بهروز مینایی بیدگلی

دانشیار، گروه هوش مصنوعی، دانشکده کامپیوتر دانشگاه علم و صنعت ایران، تهران، ایران. رایانامه: b_minai@iust.ac.ir

چکیده

هدف: عوامل جوئی متعددی بر تعداد بیماران بیمارستان‌ها اثرگذار است که بدون در نظر گرفتن آنها از امکانات و نیروی انسانی بیمارستان‌ها، استفاده غیربهبینه خواهد شد. هدف از اجرای این پژوهش، کشف دانش پنهان بین عوامل جوی با تعداد بیماران بیمارستان با استفاده از داده‌کاوی است.

روش: در این پژوهش، ارتباط بین عوامل جوئی با تعداد بیماران بیمارستان تخصصی کودکان دکتر شیخ مشهد با استفاده از رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی بررسی می‌شود. بدین منظور، پس از آمارگیری جداگانه از تعداد بیماران بخش‌های نفرولوژی، هماتولوژی، اورژانس و پی‌سی‌یو این بیمارستان، ارتباط بین تعداد این بیماران با عوامل جوئی شامل دمای هوا، رطوبت نسبی هوا، سرعت باد، فشار هوا و آلودگی هوا تحلیل شده است. داده‌های این پژوهش مربوط به ۱۹ ماه است و با مراجعه به اسناد و مدارک به‌دست آمده است. برای انتخاب ویژگی، تمام زیرمجموعه‌های عوامل جوئی، جست‌وجو شده و اثر تمام آنها بر تعداد بیماران با استفاده از رگرسیون خطی ارزیابی شده است. همچنین برای کاوش قواعد، از رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی که بر اساس الگوریتم شناخته شده آپریوری است، استفاده شده است.

یافته‌ها: نتایج به‌دست‌آمده گویای قواعدی است که ارتباط بین تعداد متفاوت بیماران در بخش‌های مختلف بیمارستان را با تغییر عوامل جوئی نشان می‌دهد.

نتیجه‌گیری: با استفاده از روش‌های به‌کار برده شده در این پژوهش، می‌توان ارتباط بین عوامل جوئی و تعداد بیماران بیمارستان را بررسی کرد. همچنین قواعد به‌دست‌آمده به مدیران کمک می‌کند که برای منابع بیمارستان با توجه به تعداد متفاوت بیماران برنامه‌ریزی بهینه‌ای انجام دهند.

کلیدواژه‌ها: رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی، الگوریتم آپریوری، رگرسیون خطی، عوامل جوئی، تعداد بیماران بیمارستان.

استناد: حدادیان، سیم‌ا؛ ناجی عظیمی، زهرا؛ مطهری فریمانی، ناصر؛ مینایی بیدگلی، بهروز (۱۳۹۸). کاوش قواعد ارتباط‌دهنده عوامل جوئی با تعداد بیماران بیمارستان با استفاده از رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی. مدیریت صنعتی، ۱۱(۴)، ۵۷۵-۵۹۹.

مقدمه

بیمارستان‌ها از مهم‌ترین سازمان‌های ارائه خدمات بهداشتی درمانی هستند (ادهم، مهدوی، مهرتک، ابراهیمی و آذری، ۱۳۹۴). به‌طور قطع، تعداد بیماران در تمام روزهای سال یکنواخت نیست و تغییر می‌کند. عوامل مختلفی بر تعداد بیماران اثرگذار است که در نظر نگرفتن این عوامل در برنامه‌ریزی برای پزشکان، نگاهی ساده‌انگارانه به استفاده بهینه از منابع انسانی و امکانات پزشکی در بیمارستان‌هاست. در این راستا، می‌توان به پژوهش‌های انجام شده محمدی، آذر و زارعی متین (۱۳۸۴)، ادهم و همکاران (۱۳۹۴) و توتی و همکاران^۱ (۲۰۱۶) اشاره کرد که تنها با توجه به یک عامل جوئی یا میانگین تعداد بیماران در سال‌های گذشته، تعداد مراجعه بیماران را پیش‌بینی کرده‌اند.

ذکر این نکته نیز مهم است که میزان تأثیرپذیری افراد از عوامل جوئی با توجه به سن آنها متفاوت است و تخمین تعداد بیماران نباید برای تمام دوره‌های عمر افراد یکسان در نظر گرفته شود (توتی و همکاران، ۲۰۱۶). بیمارستان‌ها و مراکز بهداشتی، از جمله سازمان‌هایی هستند که با حجم زیاد داده و اطلاعات مواجه‌اند. از طرفی داده‌کاوی روشی است که به‌منظور کشف و استخراج اطلاعات از پایگاه داده‌ها، استفاده می‌شود. در واقع، داده‌کاوی به‌منظور جست‌وجوی اطلاعات کلی جدید، ارزشمند و غیربدیهی، از میان حجم زیاد داده‌ها به‌کار می‌رود و به این سازمان‌ها در جهت کشف اطلاعات بیشتر، کمک می‌کند (زوکو^۲، ۲۰۱۹).

هر روز، بیمارستان‌ها پذیرای بیماران مختلفی هستند که تعداد آنها از یک بخش به بخش دیگر متفاوت است. از طرفی، تعداد بیماران هر بخش نیز در روزهای مختلف سال متغیر است. تحلیل ارتباط بین تعداد بیماران بیمارستان‌ها و عوامل جوئی، نیازمند بررسی پایگاه‌های اطلاعاتی موجود است؛ اما پیدا کردن ارتباط بین عوامل متعدد از میان حجم زیاد داده‌ها، به روش کارایی نیاز دارد. در میان روش‌های داده‌کاوی، روش کاوش قواعد^۳ به‌منظور کشف ارتباطات جالب و با اهمیت بین اقلام اطلاعاتی، در پایگاه داده‌های بزرگ و انبارهای رکورد استفاده می‌شود. کاوش قواعد وابستگی^۴ یکی از تکنیک‌های اصلی داده‌کاوی و مهم‌ترین شکل از کشف و استخراج الگوهای محلی در سیستم‌های یادگیری است و روابط را به‌صورت اگر - آنگاه نشان می‌دهد (ونگ^۵، ۲۰۱۱). رده‌بندی مبتنی بر کاوش قواعد وابستگی^۶، مانند کاوش قواعد وابستگی، به‌دنبال یافتن روابط میان خصیصه‌ها و مقادیر به‌صورت اگر - آنگاه است، با این تفاوت که در قواعد رده‌بندی، قسمت نتیجه تنها شامل مقادیر یا خصیصه‌های از پیش تعیین شده، به نام رده است؛ حال آنکه چنین خصوصیتی برای قواعد وابستگی برقرار نیست.

مسئله‌ای که این پژوهش بدان پرداخته است؛ استخراج قواعد به‌منظور پیش‌بینی تعداد بیماران، بر اساس عوامل جوئی است که از آن بتوان برای بهینه‌سازی منابع استفاده کرد. نبود سیستمی مناسب برای بررسی تناسب بین تقاضا و عرضه منابع بیمارستان برای روزهای مختلف سال یا به بیان دیگر، بی‌اطلاعی از تعداد بیماران در روزهای آینده، باعث می‌شود که برنامه‌ریزی دقیقی برای عرضه منابع بیمارستان انجام نشود. این پژوهش، به‌دنبال تحلیل ارتباط و کشف دانش پنهان بین تعداد بیماران بخش‌های مختلف بیمارستان با عوامل جوئی مختلف، از جمله دمای هوا، رطوبت نسبی

1. Toti, & et al.
3. Rule mining
5. Weng

2. Zucco
4. Association rule mining
6. Classification based on association rule mining

هوا، سرعت باد، فشار هوا و آلودگی هواست که با استفاده از آن می‌توان به پیش‌بینی تعداد بیماران در بیمارستان پرداخت. تمرکز این پژوهش بر بیمارستان‌های تخصصی کودکان است و دلیل انتخاب این‌گونه بیمارستان‌ها، مراجعان کمتر از ۱۸ سال بوده است. افراد بیمار در این رده سنی، بیشتر تحت تأثیر عوامل جوئی قرار می‌گیرند و در نتیجه، تعداد بیماران در زمان‌های مختلف با توجه به عوامل جوئی مختلف تغییر می‌کند. در این پژوهش بیمارستان تخصصی کودکان دکتر شیخ مشهد برای بررسی انتخاب شده است و به‌منظور کاوش قواعد در آن، از رده‌بندی مبتنی بر کاوش قواعد وابستگی رده‌بندی چندبُعدی^۱ استفاده شده است.

از نوآوری‌های این پژوهش، می‌توان به در نظر گرفتن عامل آلودگی هوا، همراه با عوامل دیگر جوئی به‌منظور رده‌بندی و پیش‌بینی تعداد بیماران اشاره کرد. نوآوری دیگر این پژوهش، ترکیب موضوع «قواعد وابستگی چندبُعدی» با موضوع «رده‌بندی مبتنی بر قواعد وابستگی» و به سخن بهتر، معرفی موضوع «رده‌بندی مبتنی بر قواعد وابستگی چندبُعدی» است. همچنین استفاده از روش‌های ترکیبی داده‌کاوی استفاده‌شده در این پژوهش، در حوزه بیمارستان برای پیش‌بینی تعداد بیماران، نوآوری دیگر این پژوهش است.

پیشینه نظری پژوهش

در ادامه به توضیح تأثیر پیش‌بینی تعداد بیماران بر عملکرد بیمارستان، پیش‌پردازش داده‌ها^۲، انتخاب ویژگی (عوامل مؤثر)^۳ و همچنین معرفی روش کاوش قواعد، رده‌بندی مبتنی بر کاوش قواعد وابستگی، رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی و قواعد وابستگی کمیاب^۴ که در این پژوهش استفاده شده‌اند، پرداخته خواهد شد.

تأثیر پیش‌بینی تعداد بیماران بر عملکرد بیمارستان

یکی از مسائل اصلی در بیمارستان‌ها، بررسی تعادل بین تعداد بیماران با منابع بیمارستان، شامل منابع انسانی و غیرانسانی است. شناسایی و حذف بی‌تناسبی‌های بین تقاضا و عرضه منابع بیمارستان، باعث می‌شود که هزینه‌های غیرضروری کاهش و کارایی و اثربخشی افزایش یابد. پیش‌بینی عرضه منابع بیمارستان باید بر اساس پیش‌بینی تقاضا باشد، نه اینکه فقط با توجه به یک برنامه توافقی یا برنامه‌های پیشین انجام شود. زمانی مشکل بیشتر می‌شود که برنامه‌ریزی برای مدت زمان طولانی انجام شده باشد (ادهم و همکاران، ۱۳۹۴).

پیش‌بینی تعداد بیماران یا به بیان دیگر، پیش‌بینی تقاضای منابع بیمارستان به عوامل مختلفی بستگی دارد که یکی از این عوامل مهم، آب‌وهواست. در نظر گرفتن این عوامل، بینش جامع‌تری را در اختیار برنامه‌ریزان و مدیران برای برنامه‌ریزی بهینه قرار می‌دهد. برای پیش‌بینی تعداد بیماران، روش‌های مختلفی وجود دارد، مانند: روش‌های کیفی، از جمله قضاوت خبرگان و تکنیک دلفی و روش‌های کمی، از جمله الگوریتم‌های پیش‌بینی، تجزیه و تحلیل روند و شبیه‌سازی (محمدی و همکاران، ۱۳۸۴).

1. Classification based on Multidimensional association rule mining
3. Feature selection

2. Data preprocessing
4. Rare association rule

پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها، نخستین گام در داده‌کاوی است که به منظور رفع نقص‌ها، ناسازگاری‌ها و نبود بعضی داده‌ها انجام می‌شود. پیش‌پردازش داده‌ها اصولاً شامل چهار بخش است که از هر یک بنا به نیاز و ویژگی‌های داده‌های در دسترس استفاده می‌شود و عبارت‌اند از: پاکسازی داده‌ها^۱ به منظور رفع داده‌های ناسازگار، یکپارچگی داده‌ها^۲ به منظور انسجام داده‌هایی که از منابع مختلف جمع‌آوری شده‌اند، کاهش داده‌ها^۳ به منظور حذف بعضی داده‌های غیرمفید و تبدیل داده‌ها^۴ به منظور نرمال‌سازی یا گسسته‌سازی داده‌ها در صورت نیاز (نورث^۵، ۲۰۱۲: ۳۰؛ ژانگ و ژانگ^۶، ۲۰۰۷: ۴).

انتخاب ویژگی (عوامل مؤثر)

در پردازش‌های مختلف، این نتیجه به دست آمده است که لزوماً استفاده از تمام عوامل موجود به بهترین کارایی در استخراج قواعد منجر نمی‌شود. به طور معمول اگر بخشی از عوامل مرتبط به کار گرفته شوند، سرعت حل و دقت نیز افزایش می‌یابد (جاوید، ببری و سعید^۷، ۲۰۱۲). روش‌های مختلفی برای انتخاب ویژگی وجود دارد که تمام آنها از دو تابع اصلی به نام تابع تولیدکننده^۸ و تابع ارزیابی^۹ تشکیل شده‌اند. توابع تولیدکننده مختلف شامل جست‌وجوی کامل، جست‌وجوی مکاشفه‌ای و جست‌وجوی تصادفی، زیرمجموعه‌های کاندید را پیدا می‌کند و توابع ارزیابی مختلف شامل توابع مبتنی بر فاصله، مبتنی بر اطلاعات، مبتنی بر وابستگی، مبتنی بر سازگاری و مبتنی بر خطای طبقه‌بندی، زیرمجموعه مدنظر را بر اساس روش مشخص، ارزیابی می‌کند (داش و لیو^{۱۰}، ۲۰۰۳).

کاوش قواعد

یکی از روش‌های داده‌کاوی و مبتنی بر یادگیری ماشین، کاوش قواعد است که به دو رویکرد کاوش قواعد وابستگی و رده‌بندی دسته‌بندی می‌شود (ژانگ و ژانگ، ۲۰۰۷: ۸). کاوش قواعد رده‌بندی، مانند کاوش قواعد وابستگی، به دنبال کاوش ارتباط بین خصیصه‌ها یا ویژگی‌های مختلف مبتنی بر قواعد، به صورت اگر - آنگاه است. هدف کاوش قواعد وابستگی، بررسی ارتباط بین خصیصه‌هاست، در صورتی که رده‌بندی مبتنی بر کاوش قواعد وابستگی، به دنبال ساختن قواعد رده‌بندی برای پیش‌بینی داده‌های آینده است. همچنین در کاوش قواعد وابستگی، ممکن است چندین خصیصه در قسمت چپ قواعد قرار بگیرند؛ اما در رده‌بندی مبتنی بر کاوش قواعد وابستگی، فقط یک خصیصه به عنوان رده در قسمت چپ قواعد قرار می‌گیرد (ثباتح^{۱۱}، ۲۰۰۷).

رده‌بندی مبتنی بر کاوش قواعد وابستگی

در تعریف قواعد وابستگی، می‌توان این طور بیان کرد که اگر $I = \{I_1, I_2, \dots, I_n\}$ مجموعه خصیصه‌ها و n تعداد خصیصه‌ها باشد و همچنین تمامی رکوردها با $DB = \{T_1, T_2, \dots, T_m\}$ نشان داده شود، به طوری که $T_i \subseteq I$ باشد، آنگاه

1. Data cleaning
3. Data reduction
5. North
7. Javed, Babri & Saeed
9. Evaluation function
11. Thabtah

2. Data integration
4. Data transformation
6. Zahng & Zhang
8. Generation function
10. Dash & Liu

یک قاعده وابستگی به صورت $X \rightarrow Y$ تعریف می‌شود، به طوری که $X \subseteq I, Y \subseteq I$ و $X \cap Y = \emptyset$ برقرار باشد (الهامودی، سنگفنگ و الصلحی^۱، ۲۰۱۶).

حدود دو دهه است که پژوهشگران از رده‌بندی با استفاده از کاوش قواعد وابستگی استفاده می‌کنند (هادی، آبوراب و الهواری، ۲۰۱۶). اولین بار لیو، سو و ما^۲ (۱۹۹۸) کاوش قواعد رده‌بندی را با استفاده از الگوریتم آپریوری که قبل از آن در کاوش قواعد وابستگی مطرح بود، انجام دادند. رده‌بندی مبتنی بر کاوش قواعد وابستگی، مانند کاوش قواعد وابستگی، ارتباطات بین خصیصه‌ها را به صورت اگر-آنگاه نشان می‌دهد. با این تفاوت که در قواعد رده‌بندی، خصیصه‌های قسمت آنگاه قواعد از پیش تعیین شده و نشان‌دهنده یک رده است (کانو، ظفرا و وتورا^۳، ۲۰۱۳). در این صورت X شامل یک یا چند خصیصه و Y مجموعه رده‌هاست.

رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبعدی

اگر خصیصه‌های قاعده وابستگی از یک بُعد مانند خرید باشند، به آن قاعده یک‌بعدی و اگر خصیصه‌ها از چند بُعد، مانند خرید، درآمد، سن و... تشکیل شده باشند، قاعده چندبعدی نامیده می‌شوند (سریدوی و راماراج^۴، ۲۰۱۳).

قواعد وابستگی کمیاب

قواعد وابستگی کمیاب، شامل خصیصه‌های هستند که در یک پایگاه داده، کم تکرار شده‌اند؛ اما با یکدیگر ارتباط‌های قوی دارند؛ به طوری که قواعد متشکل از این نوع مجموعه خصیصه‌ها، قواعد کم‌پشتیبان، اما پُر اطمینان هستند (بت و پرتیک^۵، ۲۰۱۴؛ براه و ناث^۶، ۲۰۱۸). توجه به قواعد کمیاب در پایگاه داده‌های پزشکی، حائز اهمیت است؛ زیرا در این نوع داده‌ها، داشتن خصیصه‌های کم‌تکرار اما بااهمیت، رایج است. با در نظر گرفتن مقدار کم برای پشتیبان، می‌توان به قواعدی که شامل خصیصه‌های کمیاب‌اند، دست یافت؛ سپس میزان جذابیت و قوی بون قواعد را بر اساس معیار اطمینان سنجید (راحل، دنگمی، ویهوا و پریزو^۷، ۲۰۰۴؛ کوه و ناتان^۸، ۲۰۱۰؛ ونگ، ۲۰۱۱).

در قسمت پیشینه نظری تلاش شده است که ابتدا موضوع اصلی این پژوهش، یعنی پیش‌بینی تعداد بیماران در بیمارستان و ضرورت انجام آن توضیح داده شود. سپس مفاهیم داده‌کاوی، از جمله پیش‌پردازش داده‌ها، به‌منظور آماده‌سازی آنها برای اجرای الگوریتم کاوش قواعد تشریح شده و مفهوم کاوش قواعد که به دو دسته کاوش قواعد وابستگی و کاوش قواعد رده‌بندی تقسیم می‌شود و همچنین، مفهوم رده‌بندی مبتنی بر کاوش قواعد وابستگی، توضیح داده شود. در نهایت، مفهوم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبعدی و قواعد کمیاب که در این پژوهش بر آنها تمرکز شده است، تشریح شود.

پیشینه تجربی پژوهش

در این قسمت، پیشینه تجربی پژوهش، در دو بخش و از دو جنبه بررسی شده است. در بخش اول، عوامل جوئی مختلف

1. AL-Hamodi, Songfeng & AL-Salhi
3. Cano, Zafra & Ventura
5. Bhatt & Pratik
7. Rahal, Dongmei, Weihua & Perrizo

2. Liu, Hsu & Ma
4. Sridevi & Ramaraj
6. Borah & Nath
8. Koh & Nathan

اثرگذار بر تعداد بیماران و در بخش دوم، کاربرد الگوریتم‌های کاوش قواعد در حوزه پزشکی در مقاله‌های گذشته بررسی می‌شود.

عوامل جوئی اثرگذار بر بیماری کودکان

پژوهش‌های متعددی به بررسی عوامل جوئی اثرگذار بر بیماری‌های کودکان پرداخته‌اند که در ادامه معرفی شده‌اند. عباسی رنجبر (۱۳۸۴) بررسی خود را با هدف تعیین شیوع و شدت بیماری‌های آلرژیک در کودکان شهر رشت انجام داد. روش این پژوهش، توصیفی و جامعه آماری آن، دانش‌آموزان رده اول ابتدایی از ۶۳ مدرسه و دانش‌آموزان دوم و سوم راهنمایی از ۲۲ مدرسه بودند که به‌طور تصادفی و خوشه‌ای، از بین مدارس شهر رشت انتخاب شدند. داده‌ها با نرم‌افزار SPSS و در سطح اطمینان ۹۵ درصد تجزیه و تحلیل شدند. از مقایسه آماری به‌دست‌آمده، معلوم شد که شیوع بیماری آسم در شهر رشت افزایش یافته و دلیل آن، افزایش تماس با آلرژی‌های خانگی یا آلاینده‌های هوا و محیطی بوده است. عساری، مدرسی، حق‌جوی و لاهیجان‌زاده (۱۳۸۹) ارتباط بین آلودگی هوا، به‌ویژه ذرات معلق بر سیستم تنهائی و متعاقب آن، فاکتورهای انعقادی در کودکان را بررسی کردند. این بررسی در مقطع شش ماهه و روی ۱۱۸ کودک ۱۰ تا ۱۸ ساله در شهر اصفهان انجام شد. این افراد به روش تصادفی، از نواحی مختلف شهر اصفهان که دارای ایستگاه آلودگی هوا بودند، انتخاب شدند. با مطالعه بالینی و آزمایشگاهی و همچنین، محاسبه ضریب هم‌بستگی پیرسون، ارتباط بین آلاینده‌های محیطی و سطح ترومبومودولین و فاکتور بافتی بررسی شد. نتایج نشان داد که آلودگی هوا، به‌ویژه ذرات معلق کوچک‌تر از ده میکرون PM10 بر شاخص‌های اختلال عملکرد آندوتلیوم تأثیر دارد و این موضوع، زمینه‌ای برای شروع روندهای تنهائی و متعاقب آن، روند انعقادی و بروز آتروسکلروز از اوایل عمر است.

فرج‌زاده و حیدری (۱۳۹۱) پژوهش خود را با هدف بررسی بین وضعیت آب‌وهوا و انتشار بیماری‌های عفونی و واگیردار، انجام دادند. در این پژوهش تحلیلی - مقطعی، افراد مبتلا به گاستروانتریت، از میان کودکان زیر ۵ سال که طی سه سال به بیمارستان کودکان بندرعباس مراجعه کردند، برای نمونه انتخاب شدند. با استفاده از ضریب هم‌بستگی پیرسون، هم‌بستگی میان داده‌های هواشناسی و تعداد بیماران با علائم گاستروانتریت و با رگرسیون خطی چندمتغیره، مدل پیش‌بینی تعداد بیماران ارائه شد. نتایج نشان داد که با هر یک درجه سانتی‌گراد کاهش دما، تعداد بیماران با علائم گاستروانتریت به میزان ده نفر افزایش می‌یابد.

تالر، لیورانس، مور و لئال^۱ (۱۹۸۵) با هدف بررسی آلودگی هوا بر کودکان، پژوهشی انجام دادند. آنها داده‌های بالینی ۸۹۳ کودک مبتلا به بحران آسم را در بازه زمانی ۱۹۸۳ تا ۱۹۸۴ در یک کلینیک در اسپانیا ارزیابی کردند. با استفاده از ضریب هم‌بستگی، هم‌بستگی بین تعداد بیماران با آلودگی هوا، داده‌های هواشناسی و میزان گرده ثابت شد. همچنین با تحلیل واریانس نشان داده شد که بین بحران‌های آسم و آلودگی هوا، در کودکان زیر ۳ سال، به‌خصوص در فصل پاییز، ارتباط معناداری وجود دارد.

دسوزا، هال و بکر^۲ (۲۰۰۷) با هدف بررسی دما و رطوبت هوا بر بیماری اسهال ویروسی بین کودکان، پژوهشی

انجام دادند. نمونه پژوهش آنها کودکان زیر ۵ سال در سه شهر استرالیا بود. آنها داده‌های روزانه مربوط به دما و رطوبت هوا و تعداد بیماران کودک را طی سال‌های ۱۹۹۳ تا ۲۰۰۳ جمع‌آوری کردند و ارتباط بین این عوامل را با رگرسیون خطی سنجیدند. نتایج نشان داد که با افزایش دما و کاهش رطوبت، به‌خصوص در فصل تابستان، از میزان بیماری اسهال ویروسی کاسته می‌شود.

میرکو، وانگ، اجر، ردی و باپتیست^۱ (۲۰۰۹) اثر دما، رطوبت و فشار هوا را روی بیماری آسم در بین کودکان بررسی کردند. نمونه پژوهش آنها کودکان و نوجوانان زیر ۱۸ سال در بیمارستان میشیگان بود. آنها ۲۵۴۰۱ داده مربوط به تعداد کودکان و نوجوانان مبتلا به بیماری آسم را در بازه زمانی ۲۰۰۴ تا ۲۰۰۵ جمع‌آوری کردند و به بررسی ارتباط این عوامل با داده‌های مربوط به آب‌وهوا، از طریق تحلیل سری‌های زمانی پرداختند. نتایج نشان داد که فشار هوا با بیماری آسم ارتباط ندارد و افزایش رطوبت و دما باعث می‌شود که تعداد بیماران آسم افزایش یابد.

اونازوکا و هاشیزوما^۲ (۲۰۱۱) ارتباط دما و رطوبت هوا را با بیماری‌های عفونی کودکان بررسی کردند. نمونه پژوهش آنها بیماران کودک در ۱۲۰ مرکز درمانی در شهر فوکاگو پرفکچر^۳ ژاپن بود. داده‌های مربوط به تعداد بیماران و پارامترهای آب‌وهوایی به‌صورت روزانه، در بازه زمانی ۲۰۰۰ تا ۲۰۱۰ جمع‌آوری شده و با تحلیل‌های سری‌های زمانی، ارتباط بین آنها بررسی شده است. نتایج نشان داد که تأثیر دما و رطوبت هوا بر کودکان زیر ۱۰ سال بیشتر است و به ازای هر یک درجه سانتی‌گراد افزایش دما و ۴/۷ درصد افزایش رطوبت، تعداد بیماری‌های عفونی ۱/۲ درصد افزایش می‌یابد.

هرواز و مارکوز^۴ (۲۰۱۵) با هدف بررسی ارتباط وضعیت آب‌وهوا با بیماری آسم، پژوهشی انجام دادند. با بررسی گذشته‌نگر، پرونده پزشکی کودکان و نوجوانان ۵ تا ۱۴ ساله را طی یک دوره ۴ ساله در یکی از بیمارستان‌های اسپانیا بررسی کردند. ۳۷۱ کودک در معرض آسم شدید بودند که علائم آنها در فصل بهار و تابستان تشدید می‌شد. در نهایت، این پژوهشگران با استفاده از روش رگرسیون، مدلی برای پیش‌بینی تشدید علائم بیماری آسم با توجه وضعیت‌های آب و هوایی مختلف ارائه دادند.

کیم، کیم، جین و لی^۵ (۲۰۱۷) اثر دما، فشار و رطوبت هوا بر بیماری تشنج و صرع کودکان بررسی کردند. جامعه آماری آنان، کودکان مبتلا به بیماری صرع و تشنج در بیمارستان چانگاون^۶ کره بود. پژوهشگران ۳۴۸۳ داده را طی بازه زمانی ۲۰۰۵ تا ۲۰۱۵ جمع‌آوری کردند و با استفاده از رگرسیون غیرخطی، به بررسی ارتباط بین عوامل یادشده با تعداد بیماران پرداختند. نتایج نشان داد که فقط دما بر بیماری تشنج اثر دارد، به‌طوری که افزایش درجه دما، بیماری تشنج را کاهش می‌دهد.

مارتینا و بایلب^۷ (۲۰۱۸) ارتباط عناصر آلوده‌کننده هوا و دما را با بیماری‌های تنفسی کودکان بررسی کردند. نمونه پژوهش، کودکان بیمار یکی از مراکز بهداشتی شهر مادرید در اسپانیا بود. آنان پس از جمع‌آوری ۶۴۷۳ داده مربوط به تعداد بیماران کودک و عناصر آلوده‌کننده و دمای هوا در بازه زمانی ۲۰۱۳ تا ۲۰۱۵، با استفاده از رگرسیون چندگانه،

1. Mireku, Wang, Ager, Reddy & Baptist
 3. Fukuoka Prefecture
 5. Kim, Kim, Jin, Jin & Lee
 7. Martina & Bayleb

2. Onozuka & Hashizume
 4. Hervásal & Marcos
 6. Changhua hospital

ارتباط بین این عوامل را بررسی کردند. نتایج نشان داد که بین دی‌اکسید نیتروژن (NO_2)، منوکسید کربن (CO) و دی‌اکسید گوگرد (SO_2) با بیماری تنفسی کودکان ارتباط مثبت و بین دمای هوا با بیماری تنفس کودکان ارتباط منفی وجود دارد.

کاوش قواعد وابستگی و رده‌بندی مبتنی بر کاوش قواعد وابستگی در حوزه پزشکی

چندین پژوهش از کاوش قواعد وابستگی و رده‌بندی مبتنی بر کاوش قواعد وابستگی در حوزه پزشکی استفاده کرده‌اند که در ادامه به آنها اشاره می‌شود.

کریمی و کنعانی (۱۳۹۳) به بررسی ارتباط عناصر آلوده‌کننده هوا، مانند ذرات معلق کمتر از ۱۰ میکرون (PM_{10})، ذرات معلق کمتر از ۲/۵ میکرون ($\text{PM}_{2.5}$)، منوکسید کربن (CO)، دی‌اکسید گوگرد (SO_2)، دی‌اکسید نیتروژن (NO_2) و ازن (O_3) با بیماری آسم آلرژیک پرداختند. داده‌های آلودگی هوا مربوط به آذر سال ۱۳۹۲ است که در هر ساعت ثبت شده است و داده‌های بیماران نیز، به پرونده پزشکی ۱۰۰۰ نفر از مراجعه‌کنندگان در مرکز طبی کودکان تهران مربوط است. ارتباط بین داده‌ها با استفاده از کاوش قواعد وابستگی چندبُعدی بررسی شده است. قواعد نهایی نشان داده است که ذرات معلق کمتر از ۱۰ میکرون، ذرات معلق کمتر از ۲/۵ میکرون، دی‌اکسید نیتروژن و منوکسید کربن بر وقوع آسم آلرژیک مؤثرند و دی‌اکسید گوگرد و ازن بر بیماری آسم تأثیری ندارند.

شریعت‌پناهی، حبیبی، رفیعی، قندی و انوری (۱۳۹۶) با استفاده از کاوش قواعد وابستگی، ارتباط بین عوامل اثرگذار بر سکنه قلبی را بررسی کردند. جامعه آماری این پژوهش، ۱۰۴۶ بیمار است که اطلاعات آنها بین سال‌های ۱۳۹۰ تا ۱۳۹۳ از کلینیک ویژه دیابت بیمارستان امام خمینی تهران جمع‌آوری شده است. پس از پیش‌پردازش داده‌ها، ۵۷۳ داده تحلیل و با الگوریتم اف پی - گروث کاوش قواعد انجام شد. نتایج نشان داد که نداشتن بیماری فشار خون، کراتینین و فشارخون دیاستولیک در سطح طبیعی، با وجود فشار خون سیستولیک بالا، احتمال وقوع سکنه قلبی را افزایش نمی‌دهد. اردنز و همکاران^۱ (۲۰۱۱) با استفاده از رده‌بندی مبتنی بر کاوش قواعد وابستگی، ارتباط بین عوامل مختلف با بیماری قلبی را بررسی کردند. مورد مطالعه آنها بیماران مبتلا به حمله قلبی در یکی از بیمارستان‌های آمریکا بود. ۶۵۵ رکورد از تعداد بیماران بیمارستان جمع‌آوری شده و با کاوش قواعد تحلیل شدند. نتایج نشان داد که افراد غیرسیگاری، افراد با کلسترول پایین، جوانان غیرسیگاری و زنان بدون بیماری دیابت، کمتر دچار حمله قلبی می‌شوند.

اردنز (۲۰۰۶) با معرفی یک الگوریتم جدید رده‌بندی مبتنی بر کاوش قواعد وابستگی، ارتباط سن، جنسیت و سیگار کشیدن را با بیماری قلبی بررسی کرد. در این پژوهش ۶۵۱ رکورد از بیماران یک بیمارستان در آمریکا جمع‌آوری و با کاوش قواعد، بیماری قلبی پیش‌بینی شده است. نتایج نشان داد که الگوریتم پیشنهادی دقت قواعد کشف شده را افزایش می‌دهد و با استفاده از قواعد به‌دست آمده می‌توان با توجه به سن، جنسیت و میزان سیگار کشیدن نوع بیمار قلبی را پیش‌بینی کرد.

نهار، تصدق، تیکل و فوئب^۲ (۲۰۱۳) با استفاده از رده‌بندی مبتنی بر کاوش قواعد وابستگی، ارتباط بین مشخصات افراد و بیماری قلبی را بررسی کردند. داده‌ها از سایت UCI^۳ جمع‌آوری شدند. بعد از به‌کارگیری الگوریتم کاوش قواعد،

1. Ordonz

2. Nahar , Tasadduq, Tickle & Yi-Ping Phoebe

3. UCI Machine Learning Repository

نتایج نشان داد که مردان در مقایسه با زنان بیشتر دچار بیماری قلبی می‌شوند و هورمون استروژن در زنان، به حفظ سلامت قلب آنها کمک می‌کند.

ایوانسویک و همکاران^۱ (۲۰۱۵) با هدف کاوش ارتباط بین مشخصات کودکان با پوسیدگی زودرس دندان، از طریق رده‌بندی مبتنی بر کاوش قواعد وابستگی، مطالعه‌ای انجام دادند. مورد مطالعه آنها کودکان رده سنی ۱۳ تا ۷۱ ماه در یک مدرسه پیش دبستانی در صربستان است. ۳۷۱ رکورد برای کودکان وجود داشت که با روش کاوش قواعد آنها را تجزیه و تحلیل کردند. نتایج نشان داد که جنسیت مذکر، وزن کم هنگام تولد و مواد موجود در شیر مادر با پوسیدگی زودرس دندان ارتباط دارد.

توتی و همکارانش (۲۰۱۶) با هدف یافتن ارتباط آلودگی هوا و عناصر آلوده‌کننده هوا، شامل ذرات معلق کمتر از ۲/۵ میکرون (PM_{2.5})، منوکسید نیتروژن (NO)، منوکسید کربن (CO)، دی‌اکسید گوگرد (SO₂)، دی‌اکسید نیتروژن (NO₂) و ازن (O₃) با حمله‌های آسم کودکان، پژوهشی انجام دادند. ۲۰۹۵۹ داده مربوط به بیماران مبتلا به آسم از سازمان تیداس^۲ طی سال‌های ۲۰۰۲ تا ۲۰۱۲ جمع‌آوری و ارتباط بین عناصر آلودگی هوا و تعداد بیماران با کاوش قواعد وابستگی بررسی شد. قواعد نهایی نشان داد که بین عناصر آلوده‌کننده در دست بررسی، دی‌اکسید گوگرد، منوکسید نیتروژن، دی‌اکسید نیتروژن و ذرات معلق کمتر از ۲/۵ میکرون با آسم، ارتباط وجود دارد و بقیه عناصر بی‌تأثیرند.

براه و نات^۳ (۲۰۱۸) ارتباط بین عوامل اثرگذار بر بیماری سرطان سینه و بیماری قلبی و عروقی را با استفاده از کاوش قواعد بررسی کردند. ۳۶۵ داده از سایت UCI جمع‌آوری شد و با استفاده از روش اف پی - گروث، کاوش قواعد وابستگی و الگوریتم پیشنهادی به نام درخت الگویابی منفرد، داده‌ها تجزیه و تحلیل شدند. نتایج نشان داد که قواعد کشف شده از روش پیشنهادی مقاله بهتر از دو روش دیگر است. از سوی دیگر، مهم‌ترین قواعد در سرطان سینه شامل عوامل ضخامت و شکل سلولی توده و در بیماری قلب و عروقی شامل عوامل نوع نقص، ورزش و جنسیت هستند.

آلویدین، بسام و عبید^۴ (۲۰۱۸) با هدف معرفی یک الگوریتم جدید رده‌بندی مبتنی بر کاوش قواعد وابستگی، مطالعه خود را انجام دادند. نمونه پژوهش آنها، پیش‌بینی سرطان سینه با توجه به عوامل اثرگذار بود که بدین منظور، ۲۸۶ نمونه مربوط به بیماران سرطانی را از سایت UCI جمع‌آوری کردند. پس از تحلیل داده‌ها مشخص شد که دقت قواعد استخراجی با استفاده از روش پیشنهادی برای پیش‌بینی سرطان بیشتر است.

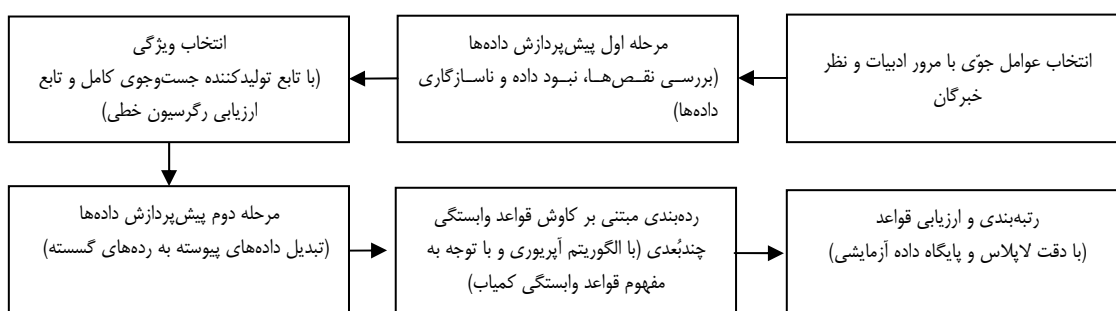
با مرور پژوهش‌های پیشین، دیده می‌شود که از کاوش قواعد وابستگی و رده‌بندی مبتنی بر کاوش قواعد وابستگی، به‌عنوان ابزارهای داده‌کاوی، در حوزه پزشکی استفاده‌های گوناگونی شده است. همچنین در برخی مقاله‌ها تأثیر عوامل جوی بر بیماری‌های مختلف بررسی شده است؛ اما آنچه در پژوهش‌های گذشته مشهود است، بررسی اثر عوامل مختلف بر تنها یک یا چند بیماری است. نتایج این تأثیرها در حوزه خاصی از پزشکی راهگشاست؛ اما برای مدیریت و برنامه‌ریزی امور بیمارستان در خصوص تعداد بیماران، بینشی در اختیار مسئولان بیمارستان قرار نمی‌دهد. از طرفی دیگر، در هر مقاله تنها به بررسی تعداد محدودی از عوامل جوی پرداخته شده است. همچنین پژوهش جامعی روی بیماری‌های کودکان و رابطه آنها با تمامی عوامل جوی اثرگذار صورت نگرفته است، از این رو، پژوهش حاضر درصدد رفع این کمبودها اجرا شده است.

1. Ivančević and et al
3. Borah, & Nath

2. Texas Emergency Department Asthma Surveillance (TEDAS)
4. Alwidian, Bassam & Obeid

روش‌شناسی پژوهش

این پژوهش از نوع کاربردی است و هدف آن یافتن قواعدی مبتنی بر ارتباط بین عوامل جوی از جمله دمای هوا، رطوبت نسبی هوا، سرعت باد، فشار هوا و آلودگی هوا با تعداد بیماران بخش‌های نفرولوژی، هماتولوژی، اورژانس و پی‌آی‌سی‌یو بیمارستان تخصصی کودکان است. بدین منظور از داده‌کاوی و به‌طور خاص برای انتخاب ویژگی، از رگرسیون خطی و برای کاوش قواعد از الگوریتم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبندی استفاده شده است. مراحل کلی کار در شکل ۱ مشاهده می‌شود.



شکل ۱. مراحل کلی اجرای پژوهش

انتخاب ویژگی (عوامل مؤثر)

همان‌طور که در بخش پیشینه نظری بیان شد، انتخاب ویژگی به دو تابع تولید و ارزیابی نیاز دارد. در روش تابع تولیدکننده و جست‌وجوی کامل در مرحله تعیین زیرمجموعه‌های نامزد، اگر تمام زیرمجموعه‌های ممکن از فضای اولیه در نظر گرفته شوند^۱، یافتن زیرمجموعه بهینه تضمین می‌شود؛ اما به‌علت پیچیدگی زمانی، این رویکرد تنها برای فضاهایی منطقی خواهد بود که تعداد عامل کمی دارند (ریف و شفیت^۲، ۲۰۱۴). از آنجا که در این پژوهش نیز عوامل محدودی وجود دارد، از این روش استفاده شده است. همچنین در تابع ارزیابی مبتنی بر وابستگی، از رگرسیون خطی استفاده شده است.

رگرسیون خطی یکی از روش‌های تحلیل رگرسیون است که علاوه بر پیش‌بینی متغیر وابسته توسط متغیرهای مستقل، شدت رابطه بین آنها را نشان می‌دهد. در رگرسیون خطی، ارتباطات بین عوامل با استفاده از یک تابع خطی نشان داده می‌شود که پارامترهای آن از داده‌ها به‌دست می‌آید (تقی‌زاده، میرشجاعیان، اصغری‌زاده و شکوری‌گنجوی، ۱۳۹۴). مدل‌های رگرسیون خطی عموماً بر اساس معیارهایی از جمله خطای جذر میانگین مربعات برآورد می‌شوند و این معیار به‌دنبال یافتن کمترین خطا بین مقدار واقعی و مقدار پیش‌بینی شده برای متغیر وابسته یا رده است. رابطه خطای جذر میانگین مربعات^۳ به‌صورت رابطه ۱ است؛ به‌طوری که γ رده واقعی، γ' رده پیش‌بینی شده و n تعداد داده‌هاست (نشاط و محلوچی، ۱۳۸۸).

۱. موسوم به رویکرد Brute force

2. Reif & Shafait

3. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_n)^2}{n}} \quad \text{رابطه (۱)}$$

با توجه به توضیحات بیان شده در بالا، می‌توان نحوه انتخاب ویژگی (عوامل مؤثر) را به صورت شکل ۲ نشان داد.

(۱) T و C_i به صورت زیر تعریف می‌شوند:
$T =$ مجموعه تمام عوامل جوئی
$C_i =$ تعداد بیماران بخش i ام و $i = 1, 2, 3, 4$.
(۲) تمام زیرمجموعه‌های مجموعه T تشکیل می‌شوند ($3^4 - 1 = 81$).
(۳) وابستگی تمام زیرمجموعه‌ها با C_i با استفاده از رگرسیون خطی محاسبه می‌شود.
(۴) در نهایت، زیرمجموعه‌ای که دارای کمترین خطای جذر میانگین مربعات است، به عنوان مجموعه ویژگی‌های نهایی انتخاب می‌شود.

شکل ۲. انتخاب ویژگی

رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی

با توجه به بررسی مفهوم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی در قسمت پیشینه نظری، مراحل این روش را می‌توان به صورت شکل ۳ توضیح داد (لیو و همکاران، ۱۹۹۸؛ هادی و همکاران، ۲۰۱۶؛ هیو، هیو، تسای، وانگ و هوانگ، ۲۰۱۶).

فرضیه‌ها و تعاریف اولیه:
۱. یک پایگاه داده به نام D وجود دارد $I = \{I_1, I_2, \dots, I_n\}$ که مجموعه همه خصیصه‌ها در D قرار می‌گیرد و Y مجموعه همه رده‌هاست.
۲. خصیصه‌های متعلق به مجموعه I از ابعاد مختلف هستند، بنابراین تمام مقادیر کمی خصیصه‌ها به مقادیر گسسته تقسیم شده و با متغیرهای کیفی نشان داده می‌شوند. هر خصیصه با m تا متغیر کیفی تعریف می‌شود. مجموعه $I' = \{(I_1, v_1), (I_1, v_2), \dots, (I_1, v_m), \dots, (I_n, v_m), \dots, (I_n, v_m)\}$ به طوری که v_i نشان‌دهنده متغیر کیفی و $1 \leq i \leq m$ باشد، در نظر گرفته می‌شود.
۳. یک خصیصه قاعده‌ای با نام $k - Ruleitem$ به صورت $\langle condet, y \rangle$ تعریف می‌شود که $condset$ نشان‌دهنده مجموعه خصیصه‌های k عضوی از مجموعه I' است و $y \in Y$.
۴. متغیر $CondSupCount$ نشان‌دهنده تعداد رکوردها در پایگاه داده D است که شامل $condset$ هستند.
۵. متغیر $RuleSupCount$ نشان‌دهنده تعداد رکوردها در پایگاه داده D است که شامل $condset$ با برچسب رده y هستند.
۶. معیار پشتیبان برابری با $100 \times \frac{RuleSupCount}{ D }$ که در آن $ D $ تعداد کل رکوردهاست.
۷. معیار اطمینان برابر است با $100 \times \frac{RuleSupCount}{CondSupCount}$.
۸. حداقل پشتیبان با $MinSupp$ و حداقل اطمینان با $MinConf$ نشان داده می‌شود.
گام‌های الگوریتم:
فاز اول:
گام (۱) تمام $k - Ruleitem$ پیدا می‌شوند.
گام (۲) برای تمامی $k - Ruleitem$ معیار پشتیبان محاسبه می‌شود.

گام ۳) اگر پشتیبان هر $Ruleitem - 1$ حداقل برابر با $MinSupp$ بود، به‌عنوان خصیصه قاعده‌ای تکراری باقی می‌ماند، در غیر این صورت حذف می‌شوند.

گام ۴) از طریق الحاق $Ruleitem - 1$ انتخاب شده در گام قبلی، مجموعه‌های $Ruleitem - 2$ تشکیل می‌شوند.

گام ۵) اگر پس از الحاق دو مجموعه تک‌خصیصه‌ای دو بُعد همسان در یک مجموعه قرار گرفتند، آن مجموعه حذف می‌شود و در غیر این صورت، معیار پشتیبان برای مجموعه حساب می‌شود.

گام ۶) اگر پشتیبان هر $Ruleitem - 2$ حداقل برابر با $MinSupp$ بود، به‌عنوان خصیصه قاعده‌ای تکراری باقی می‌ماند و در غیر این صورت حذف می‌شوند.

گام ۷) این جست‌وجوی سطحی تا پایان سطح k ادامه می‌یابد و پس از شناسایی مجموعه خصیصه قاعده‌های تکراری، فاز دو اجرا می‌شود.

فاز دوم:

گام ۱) برای تمامی مجموعه خصیصه قاعده‌های تکراری شناسایی شده در مرحله قبل، معیار اطمینان برای هر کدام از آنها محاسبه می‌شود.

گام ۲) مجموعه خصیصه قاعده‌های تکراری که حداقل برابر با $MinConf$ باشد، به‌عنوان مجموعه قواعد نهایی انتخاب می‌شود.

شکل ۳. مراحل رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی

رتبه‌بندی و ارزیابی قواعد^۱

بعد از یافتن مجموعه قواعد نهایی، باید قواعد با استفاده از پایگاه داده آزمایشی^۲ ارزیابی شوند که در این پژوهش از معیار دقت لاپلاس^۳ استفاده شده است. شایان ذکر است که این معیار زمانی که قواعد نهایی محاسبه و رتبه‌بندی شدند، به‌کار می‌رود. به همین منظور نحوه رتبه‌بندی قواعد در شکل ۴ توضیح داده شده است (ودین^۴، ۲۰۱۴). گفتنی است که رتبه‌بندی قواعد برای هر رده به‌طور جداگانه انجام می‌شود.

اگر یک مجموعه قاعده $R = \{R_1, R_2, \dots, R_n\}$ وجود داشته باشد و دو قاعده به‌صورت R_a و R_b از آن انتخاب شود، R_a به R_b ارجح است اگر:

- ۱) اطمینان R_a از اطمینان R_b بزرگ‌تر باشد؛
- ۲) اطمینان هر دو قاعده یکسان است، اما پشتیبان R_a از پشتیبان R_b بزرگ‌تر باشد؛
- ۳) اگر دو مقدار اطمینان و پشتیبان یکسان باشد، قاعده‌ای برتر است که تعداد ویژگی‌های قسمت راست آن بیشتر از دیگری باشد.

شکل ۴. رتبه‌بندی قواعد

دقت لاپلاس، نوعی روش پس - هرس^۵ است که با استفاده از رابطه ۲ برای هر قاعده محاسبه می‌شود و ضریب خطای آنها را نشان می‌دهد. در رابطه ۲، $P_c(r)$ تعداد نمونه‌های آزمایشی است که با قاعده r مطابقت دارد؛ $P_{tot}(r)$ تعداد نمونه‌های آزمایشی است که با قسمت «اگر» قاعده مطابقت دارد و m تعداد رده‌ها را نشان می‌دهد (ودین، ۲۰۱۴).

1. Rules ranking & evaluation
3. Laplas Accuracy
5. Post-Pruning

2. Testing dataset
4. Wedyan

$$Laplace(r) = \frac{(P_c(r) + 1)}{(P_{tot}(r) + m)} \quad \text{رابطه ۲}$$

یافته‌های پژوهش

در این قسمت پس از بررسی مورد مطالعه، به ارائه یافته‌های پژوهش پرداخته خواهد شد.

مورد مطالعه

بیمارستان تخصصی کودکان دکتر شیخ، بیمارستانی است که سال ۱۳۳۰ در شهرستان مشهد تأسیس شد و در حال ارائه خدمات تخصصی و فوق تخصصی به کودکان و نوزادان است. بیمارستان تخصصی کودکان دکتر شیخ مشهد، از چهار بخش جداگانه بستری به نام نفرولوژی، هماتولوژی، اورژانس و پی‌آی‌سی‌یو تشکیل شده که برای کودکان و نوجوانان زیر ۱۸ سال است.

عوامل جوئی

عوامل جوئی در نظر گرفته شده در این پژوهش، عبارت‌اند از: دمای هوا، رطوبت نسبی هوا، سرعت باد، فشار هوا و آلودگی هوا. نحوه انتخاب این عوامل بدین گونه بود که پس از مرور پژوهش‌های پیشین (جدول ۱)، تمام عوامل جوئی از مقاله‌های قبل استخراج شدند و فهرست اولیه‌ای از این عوامل تهیه شد. این فهرست در اختیار خبرگان پژوهش که همان پزشکان متخصص کودکان در بیمارستان دکتر شیخ مشهد بودند، قرار گرفت و طبق نظر آنان عوامل انتخاب شدند. بعد از انتخاب عوامل جوئی، در مرحله انتخاب ویژگی با روش رگرسیون، برای هر بخش بیمارستان این عوامل، به‌طور جداگانه انتخاب شدند و بعد از تهیه فهرست دوم، از خبرگان نظرسنجی به عمل آمد. توضیحات بیشتر در این باره، در قسمت انتخاب ویژگی مفصل بیان شده است.

جدول ۱. استخراج عوامل جوئی از پژوهش‌های پیشین

عامل جوئی	نوع بیماری	نوع تأثیر	منبع
آلودگی هوا	آسم	افزایش آلودگی هوا، افزایش بیماری را باعث می‌شود.	عباسی رنجبر، ۱۳۸۴
آلودگی هوا	آسم	افزایش آلودگی هوا، افزایش بیماری را باعث می‌شود.	عصاری، مدرسی، حق جوئی و لاهیجان‌زاده، ۱۳۸۹
آلودگی هوا	اختلال عملکرد آندوتلیوم	افزایش آلودگی هوا، افزایش بیماری را باعث می‌شود.	فرج زاده و حیدری، ۱۳۹۱
دما	گاستروانتریت	کاهش دما، افزایش بیماری را باعث می‌شود.	تالر، لیورانس، مور و لئال، ۱۹۸۵
آلودگی هوا	آسم	افزایش آلودگی هوا، افزایش بیماری را باعث می‌شود.	دسوزا، هال و بکر، ۲۰۰۷
دما و رطوبت هوا	اسهال ویروسی	افزایش دما و کاهش رطوبت، افزایش بیماری را باعث می‌شود.	میرکو، وانگ، اجر، ردی و بابتیست، ۲۰۰۹
دما، رطوبت و فشار هوا	آسم	فشار هوا با بیماری ارتباط ندارد و افزایش رطوبت و دما، افزایش بیماری را باعث می‌شود.	اونازوکا و هاشیزوما، ۲۰۱۱
دما و رطوبت هوا	بیماری‌های عفونی	افزایش دما و رطوبت، افزایش بیماری را باعث می‌شود.	کیم، جین، لی، کیم و کیم، ۲۰۱۷
دما، فشار و رطوبت هوا	تشنج و صرع	فشار و رطوبت هوا با بیماری ارتباط ندارد و افزایش دما، باعث می‌شود بیماری کاهش یابد.	مارتینا و بایلب، ۲۰۱۸

گردآوری داده‌ها

داده‌های مربوط به تعداد بیماران در طول ۱۹ ماه، از ابتدای سال ۱۳۹۶ تاکنون، از بیمارستان تخصصی کودکان دکتر شیخ مشهد در بخش‌های نفرولوژی، هماتولوژی اورژانس و پی‌آی‌سی‌یو که روزانه به‌طور جداگانه ثبت می‌شود، جمع‌آوری شدند. همچنین داده‌های مربوط به آلودگی هوا از مرکز پالایش آلاینده‌های زیست‌محیطی و دمای هوا، رطوبت هوا، سرعت باد و فشار هوا از سایت هواشناسی مشهد به‌دست آمد. گفتنی است که چهار پایگاه داده برای هر بخش بیمارستان به‌طور جداگانه در نظر گرفته شده و هر پایگاه داده، دارای ۵۸۱ رکورد و ۴۰۶۷ داده بوده است.

مرحله اول پیش‌پردازش داده‌ها

در جدول ۲ داده‌های جمع‌آوری شده درج شده است. هر سطر از جدول نشان‌دهنده یک رکورد است.

جدول ۲. مقادیر عوامل جوئی و تعداد بیماران

رکوردها	تعداد بیماران				عوامل جوئی				
	پی‌آی‌سی‌یو	نفرولوژی	هماتولوژی	اورژانس	آلودگی هوا	سرعت باد	دمای هوا	فشار هوا	رطوبت هوا
۱	۱۰	۷	۲۸	۲۸	۷۷	۱۸	۲۰	۹۰۴/۴	۲۶/۱۸
۲	۱۰	۳	۳۱	۳۱	۷۷	۳۶	۱۹	۹۰۳/۹	۲۰/۷۴
...
۵۸۱	۸	۸	۳۵	۳۵	۳۱/۶۲	۰	۲۵	۹۰۳/۷	۳۱/۶۲

داده‌های از دست رفته با استفاده از نسخه ۲۰ نرم‌افزار IBM SPSS Statistic بررسی شده است. همان‌طور که در جدول ۳ مشاهده می‌شود، فقط یک داده از دست‌رفته مربوط به عامل آلودگی هوا وجود دارد. میانگین داده‌های مربوط به آلودگی هوا برای مقدار از دست‌رفته در نظر گرفته شده است.

جدول ۳. بررسی داده‌های از دست‌رفته

تعداد رکوردها	تعداد داده‌های از دست‌رفته	مقدار جایگزین شده
۵۸۱	۱ (عامل آلودگی هوا)	۸۲/۵۶

انتخاب ویژگی (عوامل مؤثر)

انتخاب ویژگی با تابع تولید جست‌وجوی کامل و تابع ارزیابی رگرسیون خطی با استفاده از نسخه ۷/۱ نرم‌افزار RapidMiner Studio انجام شده است. داده‌ها با مقادیر پیوسته در حالت قبل از گسسته‌سازی در نظر گرفته شده‌اند. انتخاب ویژگی برای تعداد بیماران هر یک از بخش‌های بیمارستان، به‌طور جداگانه محاسبه شده است و در جدول ۴ مشاهده می‌شود. همچنین، خطای جذر میانگین مربعات مربوط به زیرمجموعه‌های عوامل جوئی که به‌عنوان ویژگی‌های نهایی انتخاب شده‌اند، در جدول ۵ درج شده است.

جدول ۴. انتخاب ویژگی برای تعداد بیماران بخش‌های مختلف بیمارستان

عوامل جوی	رطوبت هوا	فشار هوا	دمای هوا	سرعت باد	آلودگی هوا
تعداد بیماران بخش‌های بیمارستان	اورژانس	✓	×	✓	✓
	هماتولوژی	✓	×	×	✓
	نفرولوژی	✓	×	✓	×
	پی‌سی‌سی‌یو	✓	×	✓	✓

جدول ۵. مقدار RMSE برای انتخاب ویژگی بخش‌های مختلف بیمارستان

انتخاب ویژگی برای بخش‌های مختلف بیمارستان	RMSE
اورژانس	۱۰/۸۴
هماتولوژی	۴/۳۸
نفرولوژی	۳/۰۸
پی‌سی‌سی‌یو	۱/۹۹

مرحله دوم پیش‌پردازش داده‌ها

مرحله دوم پیش‌پردازش داده‌ها، به تبدیل داده‌ها مربوط می‌شود. در این مرحله به دلیل پیوستگی و چندبُعدی بودن داده‌ها، داده‌های پیوسته به دسته‌های گسسته تبدیل شده یا به بیان دیگر گسسته‌سازی انجام می‌شود. تمامی داده‌های پیوسته به سه گروه گسسته دسته‌بندی شدند که نقاط شکست آنها در جدول ۶ نشان داده شده است. برای محاسبه نقاط شکست از روش فواصل مساوی دسته‌ها (دافرتی، کوهاوی . سهامی^۱، ۱۹۹۵) از نسخه ۷/۱ نرم‌افزار RapidMiner Studio استفاده شده است.

جدول ۶. نقاط شکست عوامل جوی و تعداد بیماران

عوامل جوی و تعداد بیماران	نقاط شکست
رطوبت نسبی هوا	$[-\infty - 35/67], [35/67 - 67/83], [67/83 - \infty]$
فشار هوا	$[-\infty - 837/56], [837/56 - 177/43], [177/43 - \infty]$
دمای هوا	$[-\infty - 8/66], [8/66 - 24/33], [24/33 - \infty]$
سرعت باد	$[-\infty - 16/0.8], [16/0.8 - 33/60], [33/60 - \infty]$
آلودگی هوا	$[-\infty - 89], [89 - 141], [141 - \infty]$
تعداد بیماران اورژانس	$[-\infty - 25], [25 - 43], [43 - \infty]$
تعداد بیماران هماتولوژی	$[-\infty - 26/33], [26/33 - 35/66], [35/66 - \infty]$
تعداد بیماران نفرولوژی	$[-\infty - 8], [8 - 13], [13 - \infty]$
تعداد بیماران پی‌سی‌سی‌یو	$[-\infty - 4], [4 - 8], [8 - \infty]$

رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبعدی

بعد از تعیین نقاط شکست برای عوامل جوئی و تعداد بیماران با مقادیر پیوسته در قسمت پیش‌پردازش داده‌ها، در این بخش عوامل جوئی، به‌جز عامل فشار هوا که در قسمت قبل حذف شد و تعداد بیماران، به سه دسته «کم»، «متوسط» و «زیاد» نام‌گذاری می‌شوند. در جدول ۷، اسامی دسته‌های عوامل جوئی و تعداد بیماران بخش‌های بیمارستان درج شده است.

جدول ۷. اسامی دسته‌های عوامل جوئی و تعداد بیماران

عوامل جوئی و تعداد بیماران	نقاط شکست
دسته‌های رطوبت هوا (H)	$[-\infty - 35/67 - \infty]$ کم، $[35/67 - 67/83 - 67/83]$ متوسط، $[67/83 - \infty - \infty]$ زیاد
دسته‌های دمای هوا (T)	$[-\infty - 8/66 - \infty]$ کم، $[8/66 - 24/33 - 24/33]$ متوسط، $[24/33 - \infty - \infty]$ زیاد
دسته‌های سرعت باد (W)	$[-\infty - 16/08 - \infty]$ کم، $[16/08 - 33/60 - 33/60]$ متوسط، $[33/60 - \infty - \infty]$ زیاد
دسته‌های آلودگی هوا (PO)	$[-\infty - 89 - \infty]$ کم، $[89 - 141 - 141]$ متوسط، $[141 - \infty - \infty]$ زیاد
دسته‌های بیماران اورژانس (N _E)	$[-\infty - 25 - \infty]$ کم، $[25 - 43 - 43]$ متوسط، $[43 - \infty - \infty]$ زیاد
دسته‌های بیماران هماتولوژی (N _H)	$[-\infty - 26 - \infty]$ کم، $[26 - 36 - 36]$ متوسط، $[36 - \infty - \infty]$ زیاد
دسته‌های بیماران نفرولوژی (N _N)	$[-\infty - 8 - \infty]$ کم، $[8 - 13 - 13]$ متوسط، $[13 - \infty - \infty]$ زیاد
دسته‌های بیماران پی‌سی‌پی (N _P)	$[-\infty - 4 - \infty]$ کم، $[4 - 8 - 8]$ متوسط، $[8 - \infty - \infty]$ زیاد

تمامی پایگاه داده‌ها با استفاده از نسخه ۱۸ نرم‌افزار IBM SPSS Modeler به دو پایگاه داده آموزشی (۷۰ درصد داده‌ها) و پایگاه داده آزمایشی (۳۰ درصد داده‌ها)، تبدیل شدند. پایگاه‌های داده آموزشی برای اجرای الگوریتم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبعدی، وارد نرم‌افزار IBM SPSS Modeler شدند.

با در نظر گرفتن حداقل اطمینان ۱۰ درصد و حداقل پشتیبان ۲ درصد بعد از رتبه‌بندی قواعد، نمونه‌ای از قواعد در جدول‌های ۸ تا ۱۱ درج شده است. تعداد قواعد کشف‌شده برای تعداد بیماران بخش اورژانس، هماتولوژی، نفرولوژی و پی‌سی‌پی، به ترتیب ۲۴۰، ۲۳، ۹۵ و ۱۷۷ است. بعد از رتبه‌بندی قواعد برای ارزیابی آنها، با استفاده از پایگاه داده آزمایشی، معیار دقت لاپلاس برای نمونه قواعد نهایی محاسبه شد که نتایج آن در جدول‌های زیر درج شده است. قواعدی که دقت بیشتری داشتند، در مقایسه با قواعد دیگر با کیفیت‌تر بودند و برای رده‌بندی و پیش‌بینی داده‌های آتی به آنها قابل اتکا هستند.

جدول ۸. قواعد نهایی تعداد بیماران بخش اورژانس

دقت (درصد)	اطمینان (درصد)	پشتیبان (درصد)	اگر - آنگاه
۵۰	۱۰۰	۲/۴۶	۱) اگر رطوبت نسبی هوا زیاد و سرعت باد متوسط باشد، آنگاه تعداد بیماران متوسط است. $if H \in (67.83, \infty), W \in (16.80 - 33.60] \rightarrow N_E \in (25 - 43]$
۴۰	۱۰۰	۲/۲۲	۲) اگر رطوبت نسبی هوا زیاد و سرعت باد متوسط، آلودگی هوا کم باشد، آنگاه تعداد بیماران متوسط است. $if H \in (67.83, \infty), W \in (16.80 - 33.60], PO \in (-\infty - 89] \rightarrow N_E \in (25 - 43]$
...

ادامه جدول ۸

دقت (درصد)	اطمینان (درصد)	پشتیبان (درصد)	اگر - آنگاه
۶۶/۶۶	۷۲/۷۲	۲/۷۱	۱۳) اگر رطوبت نسبی هوا متوسط و دمای هوا زیاد باشد، آنگاه تعداد بیماران کم است. $if H \in (35.67 - 67.83], T \in (24.33 - \infty) \rightarrow N_E \in (-\infty - 25]$
...
۳۹/۲۸	۵۶/۸۶	۱۲/۵۹	۳۶) اگر سرعت باد کم و دمای هوا زیاد و آلودگی هوا کم باشد، آنگاه تعداد بیماران کم است. $if W \in (-\infty, -16.08], T \in (24.33 - \infty), PO \in (-\infty - 89] \rightarrow N_E \in (-\infty - 25]$
...
۲۳/۰۷	۳۹/۱۳	۵/۶۷	۹۶) اگر دمای هوا متوسط و سرعت باد متوسط و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران زیاد است. $if T \in (8.66, 24.33], W \in (16.80 - 33.60], H \in (-\infty - 35.67] \rightarrow N_E \in (43 - \infty)$
...
۱۰	۳۰	۴/۹۳	۱۲۹) اگر آلودگی هوا متوسط و دمای هوا متوسط و سرعت باد متوسط باشد، آنگاه تعداد بیماران زیاد است. $if PO \in (89 - 141], T \in (8.66 - 24.33], W \in (16.80 - 33.60] \rightarrow N_E \in (43 - \infty)$
...
۱۰	۱۰	۱۴/۸۱	۲۴۰) اگر آلودگی هوا متوسط و سرعت باد کم و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران زیاد است. $if PO \in (89 - 141], W \in (-\infty, -16.08], H \in (-\infty - 35.67] \rightarrow N_E \in (43 - \infty)$

جدول ۹. قواعد نهایی تعداد بیماران بخش هماتولوژی

دقت (درصد)	اطمینان (درصد)	پشتیبان (درصد)	اگر - آنگاه
۵۰	۱۰۰	۲/۴۶	۱) اگر رطوبت نسبی هوا زیاد و سرعت باد متوسط باشد، آنگاه تعداد بیماران متوسط است. $if P \in (67.83 - \infty), W \in (16.80 - 33.60] \rightarrow N_H \in (26 - 36]$
۶۶/۶۶	۹۲	۶/۱۷	۲) اگر رطوبت نسبی هوا زیاد باشد، آنگاه تعداد بیماران متوسط است. $if P \in (67.83 - \infty) \rightarrow N_H \in (26 - 36] \in (25 - 43]$
...
۵۰	۴۱/۰۲	۳۸/۵۱	۱۲) اگر رطوبت نسبی هوا کم و سرعت باد متوسط باشد، آنگاه تعداد بیماران زیاد است. $if P \in (-\infty - 35.67], W \in (16.80 - 33.60] \rightarrow N_H \in (36 - \infty)$
۴۶/۱۵	۳۵/۳۵	۴۸/۸۸	۱۳) اگر سرعت باد متوسط باشد، آنگاه تعداد بیماران زیاد است. $if W \in (16.80 - 33.60] \rightarrow N_H \in (36 - \infty)$
...
۶۰	۱۸/۷۵	۷/۹	۱۶) اگر رطوبت نسبی هوا متوسط و سرعت باد متوسط باشد، آنگاه تعداد بیماران کم است. $if P \in (35.67 - 67.83], W \in (16.80 - 33.60] \rightarrow N_H \in (-\infty - 26)$
...
۲۱/۵۶	۱۷/۱۴	۱۷/۲۸	۱۹) اگر رطوبت نسبی هوا متوسط باشد، آنگاه تعداد بیماران کم است. $if W \in (16.80 - 33.60] \rightarrow N_H \in (-\infty - 26)$
...
۱۰	۱۰	۱۷/۲۸	۲۲) اگر رطوبت نسبی هوا زیاد و سرعت باد کم باشد، آنگاه تعداد بیماران زیاد است. $if P \in (67.83 - \infty), W \in (-\infty, -16.08] \rightarrow N_H \in (36 - \infty)$

جدول ۱۰. قواعد نهایی تعداد بیماران بخش نفرولوژی

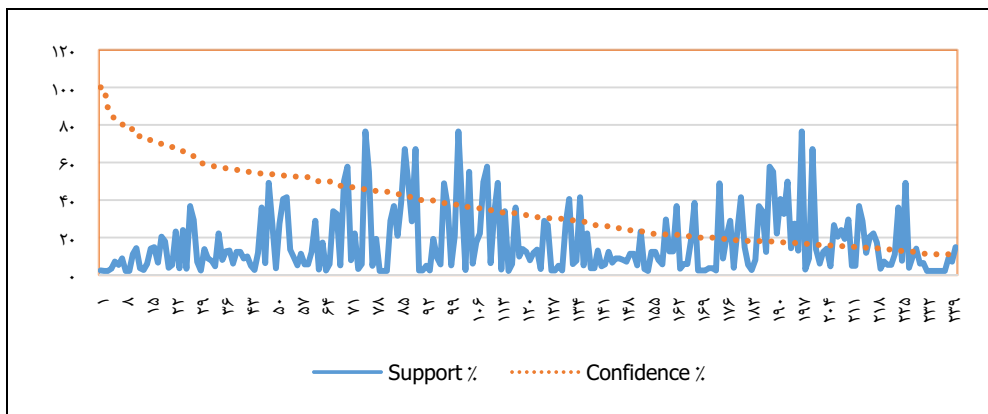
دقت (درصد)	اطمینان (درصد)	پشتیبان (درصد)	اگر - آنگاه
۵۰	۸۵	۴/۹۳	۱) اگر رطوبت نسبی هوا متوسط و آلودگی هوا متوسط باشد، آنگاه تعداد بیماران متوسط است. $if P \in (35.67 - 67.83], PO \in (89 - 141] \rightarrow N_N \in (8 - 13]$
۶۶/۶۶	۸۱/۸۱	۲/۷۱	۲) اگر رطوبت نسبی هوا متوسط و دمای هوا زیاد باشد، آنگاه تعداد بیماران متوسط است. $if P \in (35.67 - 67.83], T \in (24.33 - \infty) \rightarrow N_N \in (8 - 13]$
...
۴۷/۰۵	۶۲/۲۲	۱۱/۱۱	۶) اگر آلودگی هوا متوسط و دمای هوا متوسط و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران زیاد است. $if PO \in (89 - 141], T \in (8.66 - 24.33], P \in (-\infty - 35.67] \rightarrow N_N \in (13 - \infty)$
...
۳۰	۵۵/۵۵	۲/۲۲	۱۲) اگر دمای هوا کم و رطوبت نسبی هوا زیاد باشد، آنگاه تعداد بیماران کم است. $if T \in (-\infty - 8.66], P \in (67.83 - \infty) \rightarrow N_N \in (-\infty - 8)$
...
۳۹/۳۹	۵۴/۲۱	۲۰/۴۹	۱۷) اگر دمای هوا متوسط و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران زیاد است. $if T \in (8.66 - 24.33], P \in (-\infty - 35.67] \rightarrow N_N \in (13 - \infty)$
...
۲۲/۲۲	۴۲/۸۵	۳/۴۵	۳۶) اگر دمای هوا کم و آلودگی هوا کم باشد، آنگاه تعداد بیماران کم است. $if T \in (-\infty - 8.66], PO \in (-\infty - 89] \rightarrow N_N \in (-\infty - 8)$
...
۱۰	۱۰	۴/۹۳	۹۵) اگر رطوبت نسبی هوا متوسط و آلودگی هوا متوسط باشد، آنگاه تعداد بیماران کم است. $if P \in (35.67 - 67.83], PO \in (89 - 141] \rightarrow N_H \in (-\infty - 8)$

جدول ۱۱. قواعد نهایی تعداد بیماران بخش پی‌آی‌سی‌یو

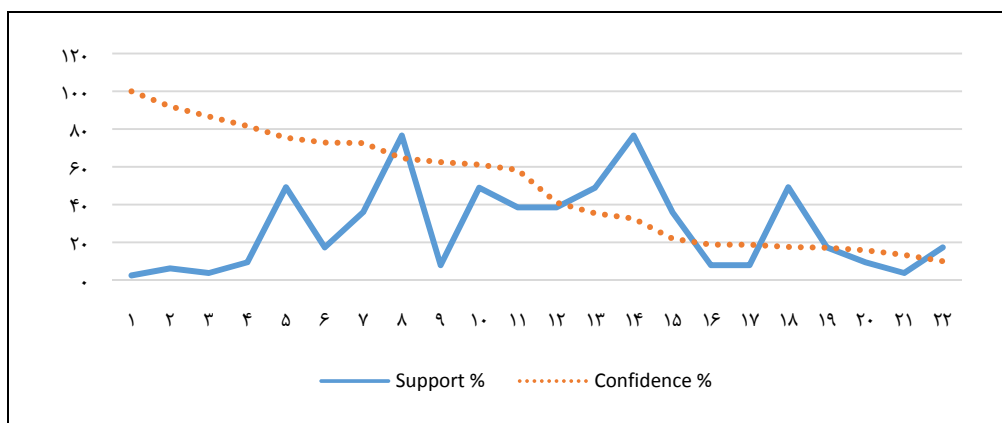
دقت (درصد)	اطمینان (درصد)	پشتیبان (درصد)	اگر - آنگاه
۷۱/۴۲	۸۲/۶۰	۵/۰۷	۱) اگر دمای هوا متوسط و سرعت باد متوسط و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران زیاد است. $if T \in (8.66 - 24.33], W \in (16.08 - 33.60], P \in (-\infty - 35.67] \rightarrow N_p \in (8 - \infty)$
۵۴/۵۴	۸۰	۲/۴۶	۲) اگر رطوبت نسبی هوا متوسط و آلودگی هوا متوسط و دمای هوا متوسط باشد، آنگاه تعداد بیماران زیاد است. $if P \in (35.67 - 67.83], PO \in (89 - 141], T \in (8.66 - 24.33] \rightarrow N_p \in (8 - \infty)$
...
۴۰	۶۳/۶۳	۲/۷۱	۳۱) اگر رطوبت نسبی هوا متوسط و دمای هوا زیاد باشد، آنگاه تعداد بیماران متوسط است. $if P \in (35.67 - 67.83], T \in (24.33 - \infty) \rightarrow N_p \in (4 - 8]$
...
۴۰	۵۸/۰۶	۷/۶۵	۴۶) اگر آلودگی هوا متوسط و سرعت باد کم و دمای هوا زیاد باشد، آنگاه تعداد بیماران متوسط است. $if PO \in (89 - 141], W \in (-\infty - 16.08], T \in (24.33 - \infty) \rightarrow N_p \in (4 - 8]$
...
۱۲/۵۰	۱۱/۱۱	۲/۲۲	۱۷۶) اگر رطوبت نسبی هوا متوسط و آلودگی هوا متوسط و سرعت باد متوسط، آنگاه تعداد بیماران کم است. $if P \in (35.67 - 67.83], PO \in (89 - 141], W \in (16.80 - 33.60] \rightarrow N_p \in (-\infty - 4)$
۱۰	۱۰/۱۸	۲۶/۶۶	۱۷۷) اگر سرعت باد متوسط و دمای هوا زیاد و آلودگی هوا کم و رطوبت نسبی هوا کم باشد، آنگاه تعداد بیماران کم است. $if W \in (16.80 - 33.60], T \in (24.33 - \infty), PO \in (-\infty - 89], P \in (-\infty - 35.67] \rightarrow N_p \in (-\infty - 4)$

بحث

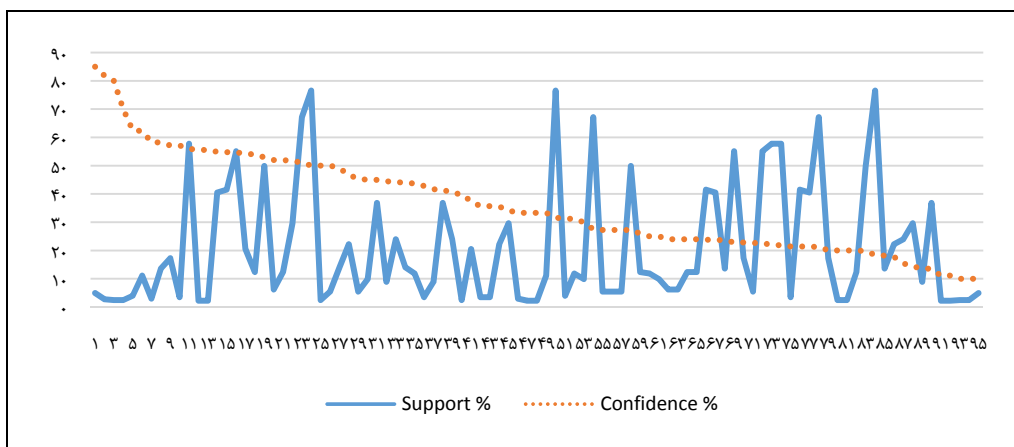
در این پژوهش، ابتدا عوامل جوئی با توجه به مرور ادبیات و نظر خبرگان جمع‌آوری شدند. سپس در قسمت پیش‌پردازش داده‌ها، اثر این عوامل بر بخش‌های بیمارستان به‌طور جداگانه بررسی شد. فشار هوا حتی بر یک بیمار از بخش‌های مختلف بیمارستان اثرگذار نبود که این نتیجه مطابق با پژوهش میرکو، وانگ، اجر، ردی و باپتیست (۲۰۰۹) است. همچنین بقیه عوامل جوئی به‌طور مختلف بر تعداد بیماران بیمارستان اثر می‌گذارد که این نتیجه با نتایج پژوهش‌های عباسی رنجبر (۱۳۸۴)، عساری و همکاران (۱۳۸۹)، فرج‌زاده و حیدری (۱۳۹۱)، تالر و همکاران (۱۹۸۵)، دسوزا و همکاران (۲۰۰۷)، میرکو و همکاران (۲۰۰۹) و کیم و همکاران (۲۰۱۷) همسو است، با این تفاوت که در این پژوهش‌ها اثر بعضی از عوامل جوئی بررسی شده است؛ اما در این پژوهش تمام عوامل جوئی در نظر گرفته شده است. در نهایت در این پژوهش با در نظر گرفتن حد آستانه‌ای برای حداقل پشتیبان و حداقل اطمینان، قواعد زیادی کشف و رتبه‌بندی شدند. مقادیر پشتیبان و اطمینان برای تعداد بیماران چهار بخش بیمارستان در شکل‌های ۵ تا ۸ نشان داده شده است. در تمام شکل‌ها، محور افقی تعداد قواعد یافت‌شده و محور عمودی میزان اطمینان و پشتیبان هر قاعده را نشان می‌دهد.



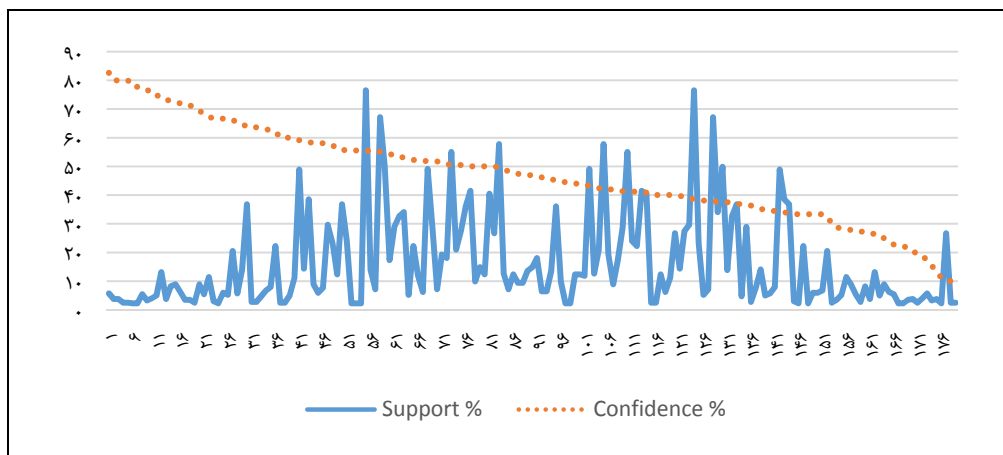
شکل ۵. نمودار مقادیر پشتیبان و اطمینان قواعد تعداد بیماران بخش اورژانس



شکل ۶. نمودار مقادیر پشتیبان و اطمینان قواعد تعداد بیماران بخش هماتولوژی



شکل ۷. نمودار مقادیر پشتیبان و اطمینان قواعد تعداد بیماران بخش نفرولوژی



شکل ۸. نمودار مقادیر پشتیبان و اطمینان قواعد تعداد بیماران بخش پی‌ای‌سی‌یو

همان‌طور که در قسمت‌های پیشین این پژوهش بیان شد، از قواعد به‌دست‌آمده می‌توان برای پیش‌بینی تعداد بیماران در روزهای آینده استفاده کرد. به همین منظور در جدول ۱۲ پیش‌بینی مقادیر عوامل جوئی مربوط به سه روز نشان داده شده است. در این روزها تعداد بیماران مشخص نیست که با استفاده از قواعد به‌دست‌آمده در این پژوهش، تعداد بیماران برای هر بخش بیمارستان پیش‌بینی می‌شود.

جدول ۱۲. مقادیر عوامل جوئی در سه روز

روز	عوامل جوئی								
	آلودگی هوا	رطوبت نسبی هوا	فشار هوا	سرعت باد	دمای هوا	اورژانس	هماتولوژی	نفرولوژی	پی‌ای‌سی‌یو
۱	۷۷	۶۲/۴۸	۹۰۲/۱	۱۸	۱۳	؟	؟	؟	؟
۲	۸۸	۱۰۰	۹۰۶/۹	۰	۳	؟	؟	؟	؟
۳	۴۰	۸۶/۸	۹۰۶/۸	۱۰/۸	۴	؟	؟	؟	؟

روش کار بدین صورت است که ابتدا مقادیر عددی عوامل جوئی، به دسته‌های گسسته با اسامی کم، متوسط و زیاد طبقه‌بندی می‌شوند، سپس این عوامل جوئی با قواعدی که رتبه‌بندی شدند، به ترتیب مقایسه خواهند شد. اولین قاعده‌ای که قسمت «اگر» آن مقادیر عوامل جوئی را پوشش داد، علامت‌گذاری می‌شود و بخش «آنگاه» این قاعده، به‌عنوان تعداد بیماران در روز بررسی شده انتخاب می‌شود. در جدول ۱۳ پیش‌بینی تعداد بیماران نشان داده شده است.

جدول ۱۳. پیش‌بینی تعداد بیماران در سه روز

روز	تعداد بیماران بخش‌های مختلف بیمارستان								
	آلودگی هوا	رطوبت نسبی هوا	فشار هوا	سرعت باد	دمای هوا	اورژانس	هماتولوژی	نفرولوژی	پی‌آی‌سی‌یو
۱	کم	زیاد	زیاد	کم	کم	متوسط	متوسط	کم	زیاد
						[۲۵-۴۳] نفر	[۲۶-۳۶] نفر	[۰-۸] نفر	(۸-۰) نفر
۲	کم	زیاد	زیاد	کم	کم	متوسط	متوسط	کم	زیاد
						[۲۵-۴۳] نفر	[۲۶-۳۶] نفر	[۰-۸] نفر	(۸-۰) نفر
۳	کم	متوسط	زیاد	کم	متوسط	متوسط	متوسط	متوسط	زیاد
						[۲۵-۴۳] نفر	[۲۶-۳۶] نفر	(۸-۱۳) نفر	(۸-۰) نفر

نتیجه‌گیری و پیشنهادها

در این پژوهش برای تحلیل ارتباط بین تعداد بیماران بیمارستان تخصصی کودکان، از داده‌کاوی و به‌طور خاص از رگرسیون خطی برای انتخاب ویژگی و از الگوریتم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی، به‌منظور کاوش قواعد استفاده شد. الگوریتم رده‌بندی مبتنی بر کاوش قواعد وابستگی چندبُعدی برای یافتن ارتباط بین دمای هوا، رطوبت نسبی هوا، سرعت باد، فشار هوا و آلودگی هوا با تعداد بیماران بخش‌های مختلف بیمارستان تخصصی کودکان دکتر شیخ مشهد از جمله اورژانس، هماتولوژی، نفرولوژی و پی‌آی‌سی‌یو به‌کار برده شد. نتایج به‌دست آمده نشان داد که تعداد بیماران بخش اورژانس با فشار هوا بی‌ارتباط و با دمای هوا، رطوبت نسبی هوا، سرعت باد و آلودگی هوا مرتبط است. تعداد بیماران هماتولوژی با فشار هوا و دمای هوا بی‌ارتباط و با رطوبت نسبی هوا، سرعت باد و آلودگی هوا مرتبط است. تعداد بیماران نفرولوژی با فشار هوا و سرعت باد بی‌ارتباط است و با دمای هوا، رطوبت نسبی هوا و آلودگی هوا ارتباط دارد و تعداد بیماران بخش نفرولوژی با فشار هوا بی‌ارتباط و با رطوبت نسبی هوا، دمای هوا، سرعت باد و آلودگی هوا مرتبط است. همچنین در نهایت تعدادی قواعد به‌دست آمد که در این قواعد سطوح مختلفی از عوامل جوئی مرتبط، تعداد بیماران را افزایش یا کاهش می‌دهد.

از کاربردهای اصلی پژوهش حاضر این است که با استفاده از قواعد، می‌توان تعداد بیماران را در بخش‌های مختلف بیمارستان و در روزهای مختلف با توجه به عوامل جوئی پیش‌بینی کرد. پیش‌بینی تعداد بیماران به بررسی و حذف متناسب‌نبودن تقاضا و عرضه منابع انسانی و غیرانسانی بیمارستان بسیار کمک می‌کند و این مسئله به تخصیص مناسب امکانات بیمارستان و نیروهای انسانی، برنامه‌ریزی دقیق‌تر برای زمان حضور پزشکان و پرستاران، کاهش هزینه‌های غیرضروری و کاهش زمان معطلی بیماران منجر می‌شود.

پیشنهاد می‌شود که در پژوهش‌های آتی، نتایج این الگوریتم با الگوریتم‌های دیگر رده‌بندی، مانند الگوریتم‌های

رده‌بندی درخت تصمیم C4/5، درخت تصمیم CART، رده‌بندی‌های ترکیبی، رده‌بندی جنگل تصادفی و... مقایسه شود. همچنین پیشنهاد می‌شود که از روش‌های مختلف گسسته‌سازی داده‌ها یا گسسته‌سازی فازی برای تحلیل ارتباط عوامل جوئی با تعداد بیماران استفاده شود.

منابع

- ادهم، داود؛ مهدوی؛ عبدالله؛ مهرتک، محمد؛ ابراهیمی، کمال؛ آذری، آرزو (۱۳۹۴). مقایسه تخصیص منابع انسانی بیمارستان‌های عمومی دانشگاهی شهرستان‌های استان آذربایجان شرقی با استاندارد کشوری. *مجله سلامت و بهداشت*، ۶(۵)، ۵۰۷-۵۱۶.
- تقی‌زاده یزدی، محمد رضا؛ میرشجاعیان حسینی، حسین؛ اصغری زاده، عزت‌اله.؛ شکوری گنجوی، حامد (۱۳۹۴). مقایسه عملکرد روش‌های رگرسیون آماری و فازی در تخمین تابع تقاضای بنزین (مطالعه موردی در ایران). *مدیریت صنعتی*، ۷(۱)، ۱-۱۹.
- شریعت پناهی، سید پیام؛ حبیبی، دانیال؛ رفیعی، محمد؛ قندی، یزدان؛ انوری، مهدی (۱۳۹۶). تعیین ارتباط بین فشار خون سیستولیک و سکت قلبی در بیماران دیابت نوع دو با استفاده از قوانین انجمنی. *مجله علمی پژوهشی دانشگاه علوم پزشکی اراک*، ۲۰(۱۲)، ۴۴-۵۰.
- عباسی رنجبر، زهرا (۱۳۸۴). شیوع علائم بیماری آسم در کودکان. *مجله علوم پزشکی دانشگاه گیلان*، ۱۴(۵۶)، ۱-۹.
- عصاری، راحله؛ مدرسی، محمدرضا؛ حق‌جوی، شقایق؛ لاهیجان زاده، احمدرضا (۱۳۸۹). بررسی ارتباط سطح آلودگی هوا با سطح سرمی ترومبومودلین و فاکتور بافتی در نوجوانان منتخب ۱۰-۱۸ ساله اصفهان. *مجله دانشکده پزشکی اصفهان*، ۲۸(۱۰۹)، ۴۲۵-۴۳۶.
- فرج‌زاده، منوچهر؛ حیدری، آریتا (۱۳۹۱). تحلیل اثر اقلیم بر بیماری‌های عفونی کودکان در منطقه بندرعباس. *مجله پژوهشی حکیم*، ۱۶(۱)، ۷۲-۷۹.
- کریمی‌پور، فرید؛ کنعانی سادات، یوسف (۱۳۹۳). بررسی وابستگی وقوع بیماری آسم آلرژیک با ویژگی‌های محیطی با استفاده از کاوش قوانین وابستگی مکانی فازی. *نشریه علمی - پژوهشی علوم و فنون نقشه‌برداری*، ۴(۳)، ۱۱۷-۱۲۹.
- محمدی، جیران؛ آذر، عادل؛ زارعی متین، حسن (۱۳۸۴). طراحی مدل برنامه‌ریزی نیروی انسانی برای بیمارستان‌های آموزشی مورد مطالعه: بیمارستان‌های آموزشی شهر اهواز. *فصلنامه علمی - پژوهشی دانشگاه شاهد*، ۱۲(۱۱)، ۷۹-۹۰.
- نشاط، نجمه؛ محلوجی، هاشم (۱۳۸۸). کنترل پیش‌بینانه کیفیت با استفاده از شبکه‌های عصبی مصنوعی (ANNs) و روش ترکیبی تحلیل رگرسیون و ANNs. *مدیریت صنعتی*، ۱(۲)، ۱۵۳-۱۷۰.

References

- Abbasi Ranjbar, Z. (2006). Prevalence of Asthma Symptoms in Children. *Journal of guilan university medical science*, 14 (56), 1-9. (in Persian)
- Adham, D., Mahdavi, A., Mehrtak, M., Ebrahimi, K., & Azari, A. (2016). Assessment of Human Resource Allocation in General University Hospitals in the Cities of East Azerbaijan Province. *Journal of health*, 6 (5), 507-516. (in Persian)

- AL-Hamodi, A., Songfeng, L., & AL-Salhi, Y. (2016). An enhanced frequent pattern growth based on mapreduce for mining association rules. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 6(2), 19-28.
- Alwidian, J., Bassam, H., & Obeid, N. (2018). WCBA: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing*, 62, 536–549.
- Assari, R., Modarresi, M., Haghjou Javanmard, SH., Lahijan-zadeh, A., Poursafa, P., Sadeghian, B., & Kelishadi, R. (2010). Evaluation of relationship between air pollution level and serum thrombomodulin and tissue factor in selected sample from 10 to 18 years old adolescents in Isfahan. *Journal of Isfahan medical science*, 28 (109), 425-436. (in Persian)
- Bhatt, U., & Pratik, P. (2014). A Recent Overview: Rare Association Rule Mining. *International Journal of Computer Applications*, 107(8), 1-4.
- Borah, A., & Nath, B. (2018). Identifying Risk Factors for Adverse Diseases using Dynamic Rare Association Rule Mining. *Expert systems with applications*, 113, 1-62.
- Cano, A., Zafra, A., & Ventura, S. (2013). An interpretable classification rule mining algorithm. *Information Sciences*, 240, 1-20.
- D'souza, R. M., Hall, G., & Becker, N. G. (2007). Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. *Epidemiol*, 136, 56-64.
- Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence*, 151(1), 155-176.
- Dougherty, J., Kohavi, V., & Sahami, M. (1995). Supervised and Unsupervised Discretization of Continuous Features. *Proceedings of the Twelfth International Conference on Machine Learning*. Tahoe City, California, USA.
- Farajzadeh, M., & Hydari, A. (2013). A. Relationship between Climate changes with Children's Infectious Diseases in Bandar Abbas, Iran. *Hakim health system research journal*, 16 (1), 72-79. (in Persian)
- Hadi, W., Aburub, F., & Alhawari, S. (2016). A new fast associative classification algorithm for detecting phishing websites. *Applied Soft Computing*, 48, 729–734.
- Hervásal, A., & Marcos, G. (2015). Can meteorological factors forecast asthma exacerbation in a paediatric population? *Allergologia et Immunopathologia*, 43 (1), 32-36.
- Hu, L., Hu, Y., Tsai, C., Wang, J., & Huang, M. (2016). Building an associative classifier with multiple minimum supports. *Speringer plus*, 5, 1-19.
- Ivančević, V., Tůsek, I., Jasmina, J., Knežević, M., Elheshk, S., & Lukovi, I. (2015). Using association rule mining to identify riskfactors for early childhood caries. *Computer Methods and Programs in Biomedicine*, 122(2), 175-181.
- Javed, K., Babri, A., & Saeed, M. (2012). Feature Selection Based on Class-Dependent Densities for High-Dimensional Binary Data. *IEEE Transactions on Knowledge and Data Engineering*, 24(3), 465-477.

- Karimipour, F., & Kananisadat, Y. (2015). Investigating the Relation between Prevalence of Asthmatic Allergy with the Characteristics of the Environment Using Fuzzy Association Rule Mining. *Journal of Geomatics Science and Technology*, 4 (3), 117-130. (in Persian)
- Kim, SH., Kim, JS., Jin, MH., Jin, C., & Lee, JH. (2017). The effects of weather on pediatric seizure: A single-center retrospective study (2005–2015). *life-sciences literature*, 609, 535-540.
- Koh, Y. & Nathan, R. (2010). *Rare Association Rule Mining and Knowledge Discovery: Technologies for infrequent and critical event detection*, New Zland, Information Science Reference.
- Liu, B., Hsu, W., & Ma, Y. (1998). Integrating Classification and Association Rule Mining, *KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, USA.
- Martina, MR. & Bayleb, MS. (2018). Impact of air pollution in paediatric consultations in Primary Health Care. *Ecological study, Pollution and paediatric consultations in Primary Health Care*, 89 (2), 80-85.
- Mireku, N., Wang, Y., Ager, J., Reddy, R.C., & Baptist, AP. (2009). Changes in weather and the effects on pediatric asthma exacerbations. *Annals of Allergy, Asthma & Immunology*, 103: 221-224.
- Mohammadi, J., Azar, A., & Zareei Matin, H. (2005). Designing a Model of Manpower Planning for Educational Hospitals in Ahwaz. *Daneshvar Raftar*, 12 (11), 79-90. (in Persian)
- Nahar, J., Tasadduq, I., Tickle, k., Yi-Ping Phoebe, Ch. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 40, 1086–1093.
- Neshat, N., & Mahlooji, H. (2010). Predictive Process Control Using Artificial Neural Networks (ANNs) and A Combined Method of Regression Analysis and ANNs. *Journal of Industrial Management*, 1 (2), 153-170. (in Persian)
- North, M. (2012). *Data Mining for the Masses*. Washington, Global Text Project.
- Onozuka, D., & Hashizume, M. (2011). The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan Science of the Total Environment. *Science of the Total Environment*, 410, 119-125.
- Ordonez, C. (2006). Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 334- 343.
- Ordonez, C., Omiecinski, E., de Braal, L., Santana, C., Ezquerro, N., Taboada, J., Cooke, D., rawczynska, E., & Garcia, E. (2001). Mining Constrained Association Rules to Predict Heart Disease. *Proceedings 2001 IEEE International Conference on Data Mining*. San Jose, USA.
- Rahal, I., Dongmei, R., Weihua, W., & Perrizo, W. (2004). Mining Confident Minimal Rules with Fixed-Consequents. In *Proceedings OF THE 16TH IEEE International Conference ON Tools With Artificial Intelligence*. IEEE Computer Society. Washington.

- Reif, M., & Shafait, F. (2014). Efficient feature size reduction via predictive forward selection. *Pattern Recognition*, 47(4):1664-1673.
- Shariatpanahi, S P., Habibi, D., Rafiei, M., Ghandi, Y., & Anvari, M. (2018). Determination of Relations between Systolic Blood Pressure and Heart Attack in Patients with Type 2 Diabetes with Association Rules. *Journal of Arak University Medical Science*, 20 (12), 44-50. (in Persian)
- Sridevi, R., & Ramaraj, E. (2013). A General Survey on Multidimensional And Quantitative Association Rule Mining Algorithms. *International Journal of Engineering Research and Applications*, 3(4), 1442-1448.
- Taghizadeh Yazdi, M., Mirshojaeian Hosseini, H., Asgharizadeh, E., & SHakouri Ganjavi, H. (2015). Relative Performance of Statistical and Fuzzy Regression Models in Estimation of Gasoline Demand in Iran. *Journal of Industrial Management*, 7 (1), 1-19. (in Persian)
- Tauler, E., Llorens-Terol, J., Mur, A., & Leal, C. (1985). Asthma and Environmental Factors. *Pediatric Research* 19, 1120-1129.
- Thabtah, F. (2007). A review of associative classification mining. *The Knowledge Engineering Review*, 22(1), 37-65.
- Toti, G., Vilalta, V., Lindner, P., Lefer, B., Macias, C., & Price, D. (2016). Analysis of correlation between pediatric asthma exacerbation and exposure to pollutant mixtures with association rule mining. *Artificial Intelligence in Medicine*, 74, 44-52.
- Wedyan, S. (2014). Review and Comparison of Associative Classification Data Mining Approaches. *International Journal of Industrial and Manufacturing Engineering*, 8(1), 34-45.
- Weng, Ch. (2011). Mining fuzzy specific rare itemsets for education data. *Knowledge-Based Systems*, 24, 697-708.
- Weng, Ch. (2016). Identifying association rules of specific later-marketed products, *Applied Soft Computing*, 38, 528-529.
- Zahng, Ch. & Zhang, SH. (2007). *Association Rule Mining Models and Algorithms*. Springer-Verlag Berlin Heidelberg New York.
- Zucco, Ch. (2019). Data Mining in Bioinformatics. *Reference Module in Life Sciences*, 1, 328-335.