

Analysis of factors affecting the length of stay of patients using clustering and association rules (Case study: Amir al-Momenin Hospital, Maragheh)

Mahdi Yousefi Nejad Attari¹, Karim Farajeyan² and Hossein Jaleb³

1. Corresponding Author, Associate Prof., Department of industrial engineering, Bon.C., Islamic Azad University, Bonab, Iran. Email: mahdi_108108@yahoo.com
2. Assistant Professor, Department of Mathematics, Bon.C., Islamic Azad University, Bonab, Iran. Email: karim.farajeyan@iau.ir
3. Assistant Professor, Department of Mathematics, Bon.C., Islamic Azad University, Bonab, Iran. Email: 1689597631@iau.ir

Article Info	ABSTRACT
Article type: Research Article Article history: Received 15 August 2025 Received in revised form 17 July 2025 Accepted 21 October 2025 Published online 1 December 2025 Keywords: Average length of stay, Clustering models, Data mining, Epiuri algorithm.	One of the major indicators in evaluating the performance of hospitals and their managers is the average length of stay of patients; given the importance of this indicator, the present study has examined the factors affecting the length of stay of hospitalized patients. This study was conducted with the aim of identifying the key factors affecting the length of stay of patients and providing practical solutions for improving the management of hospital beds. Data from 26,907 patients were analyzed using clustering models, clustering algorithms (K-Means) and association rules extraction (Apriori). The data consists of 10 numerical and discrete columns. The variables include 10 items, which are respectively: gender, marital status, hospitalization department, physician specialty, insurance, blood transfusion, surgery, type of discharge, age, and length of stay. The findings showed that the variables of surgery and blood transfusion have the greatest impact on the average length of stay in the hospital.
Cite this article: Yousefi Nejad Attari, M. & et al, (2025)., Analysis of factors affecting the length of stay of patients using clustering and association rules (Case study: Amir al-Momenin Hospital, Maragheh). <i>Enginiiring Management and Soft Computing</i>, 11 (2). Error! Bookmark not defined. DOI: https://doi.org/10.22091/jemsc.2025.11174.1195	
 © Attari et.al (2025) DOI: https://doi.org/10.22091/jemsc.2025.11174.1195 Publisher: University of Qom	

1) INTRODUCTION

Lack of financial and human resources are among the issues that have always faced hospital managers with difficulties in achieving their goals. Since the optimal use of these resources plays a significant role in the effectiveness of hospitals, one of the main challenges for hospitals in increasing efficiency and effectiveness is reducing the length of stay of patients in the hospital. One of the major indicators in evaluating the performance of hospitals is the average length of stay of patients, which, along with indicators such as bed occupancy rate and bed turnover or bed return interval, are considered the most important performance indicators of the hospital. Of course, in addition to these indicators, given the limited resources of hospitals, hospital managers always try to reduce the average length of stay of patients because the reasonable lowness of this indicator is one of the positive indicators for the status of patient management in hospitals. However, it should be noted that reducing patient stays based only on performance indicators and paying little attention to the clinical characteristics of patients will cause long-term harmful consequences for the health system. Very short stays will have inappropriate health outcomes, will prolong readmissions and increase the duration of patients' recovery. If the goal of reducing the length of stay of patients is only to increase the number of new admissions, not only will costs not be reduced, but it will also increase the cost of diseases.

2) Solution Method

Weka software was used for analysis. Weka software is an open-source tool for machine learning and data mining that implements various algorithms such as clustering and association rules.

In clustering analysis, a set of objects is grouped so that objects in one group (called a cluster) are more similar than in other groups (clusters). In this study, clustering was selected using the K-Means method with a cluster number of two to five, and each cluster was analysed individually to provide a broader view of choosing the best number of clusters for the hospital. Also, the Apriori technique was used to implement association rules in Weka software. The number of clusters for K-Means, the minimum confidence for Apriori are important parameters for evaluation.

for the two clusters, 41% of patients (1055 patients) are in cluster zero and 59% of patients (15852 patients) are in cluster one. The majority of both clusters are male. The mean age in cluster zero is approximately 34 and in cluster one is approximately 44. Most patients in cluster zero are single and most married patients are classified in cluster one. The mean length of stay in cluster zero is 1.259 and 1.581 days, respectively.

The output of the Cummins method, which is obtained using the Weka software for the number of clusters three, as shown in Figure 3, shows that according to this clustering, 26% of patients (6864 patients) are in cluster zero, 49% of patients (13261 patients) are in cluster one, and 25% of patients (6782 patients) are in cluster two. Most patients are classified in cluster one. In cluster zero, a larger number of patients are female, and in clusters one and two, most patients are male. The average age in cluster zero is approximately 36, in cluster one it is approximately 44, and in cluster two it is 38. Most patients who are in cluster zero are single, and most married patients are classified in clusters one and two. Most patients who have undergone surgery are classified in cluster two. The average length of stay in cluster two is the highest compared to the next two clusters, 2.42 days, and clusters zero and one are about 1.12 days.

Final cluster centroids:				
Attribute	Full Data (26907.0)	Cluster# 0 (6864.0)	1 (13261.0)	2 (6782.0)
number	13454	20057.4024	14875.7393	3990.7976
SexType	man	woman	man	man
age	40.24	36.1502	43.6147	37.7807
MaritalType	Married	Single	Married	Married
secname	emergency department	emergency department	emergency department	Mens surgery department
Speciality Name	Emergency Medicine	Emergency Medicine	Emergency Medicine	Orthopedics
bimname	ok	ok	ok	ok
khon	NOK	NOK	NOK	NOK
amal	NOK	NOK	NOK	OK
removecause	Doctors order	Doctors order	Doctors order	Doctors order
LOS	1.4496	1.1225	1.1189	2.4272

Time taken to build model (full training data) : 0.58 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	6864 (26%)
1	13261 (49%)
2	6782 (25%)

3) Results and discussion

Weka software was used for analysis. Weka software is an open-source tool for machine learning and data mining that implements various algorithms such as clustering and association rules.

Association rules are one of the main techniques of data mining and are almost the most important form of pattern discovery and extraction in learning systems. Association rules discover interesting connections between large sets of data that can help decision makers. In fact, association rules indicate conditions that occur repeatedly together in a data set. In this study, the a priori algorithm was used to extract association rules. According to Figure 6, these rules were implemented with support = 65% and Confidence = 80%. The following are among the important rules extracted by the a priori algorithm by the Weka software:

1. With 100% certainty, if a patient is admitted to the emergency department, he will not undergo surgery.
2. With 100% certainty, if a patient is admitted to the emergency department and has not received a blood transfusion, he will not undergo surgery.
3. With 100% certainty, if a patient is admitted to the emergency department, they will not receive a blood transfusion.
4. With 100% certainty, if a patient is admitted to the emergency department and does not undergo surgery, they will not receive a blood transfusion.
5. With 100% certainty, if a patient is admitted to the emergency department, they will not undergo surgery and a blood transfusion.
6. With 100% probability, a patient who has not undergone surgery and has been discharged from the hospital on doctor's orders will not receive a blood transfusion.
7. With 100% probability, a patient who has been discharged from the hospital on doctor's orders will not require a blood transfusion.
8. With 100% probability, a patient who has not undergone surgery will not receive a blood transfusion.

To develop the results of this study, it is suggested that other clustering methods such as SOM and fuzzy clustering be used, as well as other methods for optimizing the number of clusters such as genetic algorithm and particle swarm optimization.

Management suggestions:

1- Using the results obtained, patients admitted to the emergency department are not operated on and have a shorter length of stay. Therefore, hospital management should focus its efforts on managing patients with a longer length of stay in other departments.

2- Patients admitted to the emergency department do not need blood transfusions, so blood bank management should pay more attention to surgical rooms to reserve different blood groups.

The limitations of the study can be stated as the single hospital where the study was conducted, the lack of recording of data related to patient records, and the lack of accurate sensors to measure some parameters such as environmental factors.

4) Conclusion

By introducing cloud-based BSCM, this study bridges technology and healthcare to enhance BSCM efficiency. The findings demonstrate that cloud-based solutions can significantly improve e-healthcare performance, making CBSC an effective policy for modernizing BSCs worldwide.

REFERENCES

- Siddique, S.M., Tipton, K., Leas, B., Greysen, S.R., Mull, N.K., Lane-Fall, M., McShea, K. and Tsou, A.Y., (2021). Interventions to reduce hospital length of stay in high-risk populations: a systematic review. *JAMA network open*, 4(9), pp.e2125846-e2125846. doi:[10.1001/jamanetworkopen.2021.25846](https://doi.org/10.1001/jamanetworkopen.2021.25846)
- Amerioun A, Teymourzadeh E, Alijanzadeh M, Hakimzadeh SM. (2014). Research Needs Assessment and Priority Setting for Health Economics : A Mixed method Study in Iran. 16(1):23–8. <https://doi.org/10.1017/S1744133105001040>
- Hosseini-Shokouh SM, Sadeghian K, Ameryoun A, Zaboli R. (2019) Organizational factors affecting a patient's average length of stay in hospital: Systematic review. *J Mil Med*. 20(6):577–88. <http://doi.org/10.1097/BOT.0000000000000221>.
- Siskou O, Kaitelidou AED. (2009). Private expenditure and the role of private health insurance in Greece : status quo and future trends. 467–74. <https://doi.org/10.1007/s10198-009-0164-3>
- Nasiripour AA, Riyahi L, Gholamipour MA, others. (2010) The effect of the presence of full time gynecologist on length of stay among inpatient of gynecology ward in yazd social security hospital, 2008. *J Med Counc Islam Repub Iran*. 28(2):169–235. <https://doi.org/10.1016/j.phrp.2016.04.008>
- Yaya, L.H.P., Fortià, M.F., Canals, C.S. et al. (2015). Service quality assessment of public transport and the implication role of demographic characteristics. *Public Transp* 7, 409–428. <https://doi.org/10.1007/s12469-014-0099-7>
- Yaghoubi M, Agha Rahimi Z, Javadi M. (2013). Factors affecting in productivity index in Noor & Ali Asghar hospital in Esfahan base on hierarchical analysis process technique. *Heal Inf Manag*. 10(4):600. <https://doi.org/10.22122/him.v10i4.657>
- Mohammadkarim B, Jamil S, Pejman H, Seyyed MH, Mostafa N. (2011). Combining multiple indicators to assess hospital performance in Iran using the Pabon Lasso Model. *Australas Med J*. 4(4):175. <https://doi.org/10.4066/AMJ.2011.620>
- Angadi G, Chen R. (2007). Optimization of patient length of stay at Bridgepoint Hospital. BSc Thesis] Toronto, University of Toronto. <https://doi.org/10.1007/s001340050880>
- Back H, Cho M, Kim S, Hwang H, Song M, Yoo S. (2018). Analysis of length of hospital stay using electronic health records : A statistical and data mining approach. 1–16. <https://doi.org/10.1371/journal.pone.0195901.t004>
- [11] Molaei Fard, R. (2023). Provide a method to diagnose and optimize diabetes using data mining methods and firefly algorithm. *Enginiiring Management and Soft Computing*, 9 (1). 36-48 <https://doi.org/10.22091/JEMSC.2022.6575.1147>
- [12] Yang H, Chen YP. (2015). Expert Systems with Applications Data mining in lung cancer pathologic staging diagnosis : Correlation between clinical and pathology information. *Expert Syst Appl*. 42(15–16):6168–76. <https://doi.org/10.1016/j.eswa.2015.03.019>
- Tayefi M, Tajfard M, Saffar S, Hanachi P, Reza A, Esmaeily H, et al. (2017). Computer Methods and Programs in Biomedicine hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Comput Methods Programs Biomed*. 141:105–9. <https://doi.org/10.1016/j.cmpb.2017.02.001>
- Khajehali N, Alizadeh S. (2017). Extract critical factors affecting the length of hospital stay of pneumoni patient by data mining (case study: an Iranian hospital). *Artif Intell Med*. <https://doi.org/10.1016/j.artmed.2017.06.010>
- Fakhri Nazari, M. Fakhri Nazari, E. & Bakhsh, N. (2023). Presenting a method to reduce the sensitivity of incremental clustering algorithms of XML documents based on collective intelligence algorithms. *Enginiiring Management and Soft Computing*, 9 (2). 177-187. <https://doi.org/10.22091/jemsc.2024.9990.1176>
- Saeidi-Mobarakeh, Z. & Amoozad Khalili, H. (2024). Nonlinear Multi-Objective Optimization Model for Production Planning Based on Fuzzy Logic and Machine Learning. *Enginiiring Management and Soft Computing*, 10 (1). 155-189. <https://doi.org/10.22091/jemsc.2024.11186.1197>
- Ghadri N. (2016). Data mining with data algorithms in health trends. *J Sci Eng Elit*. (In Persian) <https://doi.org/10.1177/1460458216678443>
- Dol SM, Jawandhiya PM. (2023). Classification technique and its combination with clustering and association rule mining in educational data mining—A survey. *Engineering applications of artificial intelligence*. Jun 1;122:106071. <https://doi.org/10.1016/j.engappai.2023.106071>

تحلیل عوامل موثر بر مدت زمان اقامت بیماران با استفاده از خوشه‌بندی و قوانین انجمنی (مطالعه موردی: بیمارستان امیرالمومنین مراغه)

مهدی یوسفی نژاد عطاری^۱ , کریم فرجیان^۲  و حسین جالب^۳ 

۱. نویسنده مسئول، دانشیار گروه مهندسی صنایع، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران. رایانامه: mahdi_108108@yahoo.com

۲. استادیار گروه ریاضی، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران. رایانامه: karim.farajevan@iaui.ir

۳. استادیار گروه ریاضی، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران. رایانامه: 1680303864@iaui.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۳/۰۵/۲۹ تاریخ بازنگری: ۱۴۰۳/۱۲/۱۶ تاریخ پذیرش: ۱۴۰۴/۰۵/۱۳ تاریخ انتشار: ۱۴۰۴/۰۹/۱۰ کلیدواژه‌ها: الگوریتم آپیوری، داده کاوی، متوسط طول مدت اقامت، مدل‌های خوشه‌بندی.	یکی از شاخص‌های عمده در ارزیابی عملکرد بیمارستان‌ها و مدیران آنها، متوسط مدت زمان اقامت بیماران است. با توجه به اهمیت این شاخص، در مطالعه حاضر به بررسی عوامل موثر بر طول مدت اقامت بیماران بستری پرداخته شده است. این پژوهش با هدف شناسایی عوامل کلیدی تاثیرگذار بر طول مدت اقامت بیماران و ارائه راهکارهای عملی برای بهبود مدیریت تخت‌های بیمارستانی انجام شده است. داده‌های ۲۶۹۰۷ بیمار با استفاده از ارائه مدل‌های خوشه‌بندی الگوریتم‌های خوشه‌بندی (K-Means) و استخراج قوانین انجمنی (Apriori) تحلیل شدند. داده‌ها شامل ۱۰ ستون عددی و گسسته است. متغیرها شامل ۱۰ مورد که به ترتیب عبارتند از: جنسیت، وضعیت تاهل، بخش بستری، تخصص پزشک، بیمه، تزریق خون، عمل جراحی، نوع ترخیص، سن و طول مدت اقامت می‌باشد. یافته‌ها نشان داد که متغیرهای «عمل جراحی» و «تزریق خون» بیشترین تاثیر را بر متوسط طول مدت اقامت در بیمارستان دارند.

استناد: یوسفی نژاد عطاری، مهدی و همکاران. (۱۴۰۴). «تحلیل عوامل موثر بر مدت زمان اقامت بیماران با استفاده از خوشه‌بندی و قوانین انجمنی (مطالعه موردی: بیمارستان امیرالمومنین مراغه)». مدیریت مهندسی و رایانش نرم، دوره ۱۱ (۲)، صص: ۲۰-
Error! Bookmark not defined.

<https://doi.org/10.22091/jemsc.2025.11174.1195>



© عطاری و همکاران (۱۴۰۴).

ناشر: دانشگاه قم

(۱) مقدمه

کمیود منابع مالی و انسانی از جمله مسائلی است که همواره مدیران بیمارستان‌ها را در رسیدن به اهداف با مشکل مواجه کرده است. از آنجاییکه استفاده بهینه از این منابع نقشی بسزایی در اثربخشی بیمارستان‌ها دارد لذا کاهش مدت اقامت بیماران در بیمارستان، از چالش‌های اصلی بیمارستان‌ها در افزایش کارآیی و اثربخشی می‌باشد. یکی از شاخص‌های عمده در ارزیابی عملکرد بیمارستان‌ها، متوسط اقامت بیماران است. شاخص‌هایی مانند ضریب اشغال تخت و گردش تخت یا فاصله بازگردانی تخت، از مهمترین شاخص‌های عملکردی بیمارستان‌ها به‌شمار می‌آیند. البته در کنار این شاخص‌ها، با توجه به منابع محدود بیمارستان‌ها، مدیران بیمارستان‌ها همواره می‌کوشند تا متوسط اقامت بیماران را کاهش دهند زیرا پایین بودن منطقی این شاخص، یکی از نشانگرهای مثبت برای وضعیت مدیریت بیماران در بیمارستان‌ها است. باوجود این باید توجه داشت که کاهش اقامت بیماران فقط براساس شاخص‌های عملکردی و کم‌توجهی به ویژگی‌های بالینی بیماران، سبب پیامدهای درازمدت زیان‌باری برای سیستم سلامت خواهد بود. اقامت‌های خیلی کوتاه، نتایج نامناسبی برای سلامت بیماران داشته است و موارد پذیرش مجدد و نیز افزایش مدت زمان بهبود بیماران را طولانی خواهد کرد. اگر هدف از کاهش مدت اقامت بیمار، فقط افزایش در تعداد پذیرش‌های جدید باشد، نه تنها هزینه‌ها کاهش نخواهد یافت بلکه باعث افزایش هزینه بیماری‌ها نیز خواهد شد.

امروزه سطح پیشرفت خدمات بهداشتی و درمانی به‌عنوان یکی از شاخص‌های اندازه‌گیری رشد و توسعه اقتصادی-اجتماعی به‌شمار رفته و بیمارستان‌ها به‌عنوان یکی از منابع اجتماعی حیاتی و ضروری محسوب می‌شوند. بنابراین با توجه به اینکه مدیریت و چگونگی بهره‌برداری از خدمات بیمارستانی بر سطح رفاه اجتماعی تاثیرگذار بوده و بخش عظیمی از منابع ملی را مصرف می‌نماید، مدیریت بیمارستان باید به‌گونه‌ای باشد که در راستای سودرسانی به جامعه قدم بردارد [۱]. به‌همین دلیل اداره مطلوب این سازمان‌ها به‌عنوان اصلی‌ترین و بزرگترین بخش از نظام ارائه خدمات بهداشتی و درمانی همواره مورد توجه سیاستگذاران و مدیران بوده است [۲]. از طرفی به لحاظ محدودیت منابع سرمایه‌گذاری و نیز دیربازده بودن آن در بخش بهداشت و درمان، در کنار عواملی چون هزینه گزاف احداث بیمارستان، گرانی تجهیزات و کمبود نیروی انسانی ماهر و متخصص در این بخش، گسترش امکانات بیمارستانی را بسیار کند و دشوار می‌نماید. بنابراین حداکثر استفاده از منابع موجود با بهره‌گیری از راهکارهای برتر مدیریتی لازم خواهد بود [۳]. آمارها نشان می‌دهد بخش-های بستری بیمارستان، بیشترین هزینه‌بری را داشته و نگاهی اجمالی به حجم هزینه‌های بلائمر و تحمیلی ناشی از اشغال تخت غیرضروری به‌عنوان هزینه‌های سالانه تخت‌های بیمارستانی، استفاده از راهکارهای ارتقای بهره‌وری به‌منظور بهینه-سازی شاخص‌های کلیدی نظیر مدت اقامت بیمار را بیش‌ازپیش گوشزد می‌نماید [۴].

(۲) پیشینه تحقیق

استفاده بهینه از عوامل تولید از جمله تخت‌های بستری، نقش بسزایی در ارتقای کارآیی و اثربخشی بیمارستان‌ها داشته و یکی از چالش‌های اصلی بیمارستان‌ها در این خصوص، بهینه کردن مدت اقامت بیماران است [۵]. اما آنچه بیشتر از همه مدیران را در امر اداره بیمارستان‌ها موفق می‌نماید، حداکثر استفاده از حداقل امکانات برای ادامه هرچه بیشتر خدمات

بهداشتی و درمانی با کیفیت مطلوب و قابل قبول است [۶]. این موضوع در بیمارستان زمانی امکان‌پذیر خواهد بود که وجود هریک از واحدهای مستقر در بیمارستان‌ها از نظر مادی و معنوی در جهت سرعت بخشیدن به بهبود بیماران و بهینه‌نمودن مدت اقامت بیماران در بیمارستان عمل نمایند [۷]. از آنجاییکه متوسط مدت اقامت بیمار، به‌عنوان یکی از شاخص‌های کلیدی عملکرد و جهت سنجش کارایی بیمارستان در نظر گرفته می‌شود [۸] لذا از دیدگاه بیمارستانی، بهینه‌سازی آن و حذف موارد غیرضروری مدت اقامت، باعث افزایش کارایی از طریق افزایش ستانده در مقایسه با کاهش نهاده می‌شود. بررسی و تحلیل مدت اقامت در بیمارستان برای مدیریت به‌ویژه در تعیین اولویت‌ها و بهبود خدمات و تخصیص مناسب منابع مفید خواهد بود چراکه مدت اقامت بیماران تحت تاثیر عوامل گسترده بالینی و غیربالینی بوده که در مطالعات متعدد به اثبات رسیده است [۹].

مدت اقامت بیمار در بیمارستان، شاخص مهمی از کارایی مدیریت بیمارستان می‌باشد. کاهش تعداد روزهای بستری، منجر به کاهش خطر عفونت و عوارض جانبی دارو، بهبود کیفیت درمان و افزایش سود بیمارستانی با مدیریت بهتر تختخواب می‌شود. در این راستا بیک و همکاران [۱۰] در پژوهشی، متوسط طول مدت اقامت بیماران در بیمارستان را با استفاده از سوابق الکترونیکی بهداشت بررسی کرده‌اند. هدف از این پژوهش، تعیین عوامل موثر بر طول مدت بستری در بیمارستان براساس سوابق الکترونیکی سلامت به‌منظور مدیریت مدت اقامت در بیمارستان بود. از الگوریتم‌های داده کاوی، می‌توان برای تشخیص و پیشگیری بعضی از بیماری‌ها استفاده نمود. همچنین در مجموعه‌های آموزشی و آزمایشی، این الگوریتم‌ها را بکار می‌گیرند [۱۱]. داده‌های پژوهش، از بانک اطلاعاتی بیماران بستری‌شده در یک بیمارستان دانشگاهی در کره جنوبی بین ژانویه و دسامبر ۲۰۱۳ استخراج شده بودند. بیماران براساس سه دسته: تجزیه و تحلیل توصیفی و اکتشافی، تجزیه و تحلیل الگوی فرآیند با استفاده از تکنیک‌های فرآیند استخراجی و آماری، تجزیه و تحلیل و پیش‌بینی مدت اقامت در بیمارستان مورد تجزیه و تحلیل قرار گرفته بودند. بطور کلی ۵۵ درصد (۲۵۲۲۸) بیماران بستری طی ۴ روز مرخص شده بودند. بخش توانبخشی با ۱۵/۹ روز بالاترین مدت اقامت در بیمارستان را داشت. از بین تمام شرایط تشخیص داده شده، سکتة مغزی و سکتة قلبی با طولانی‌ترین مدت بستری در بیمارستان همراه بود. طیف وسیعی از متغیرها از جمله انتقال، زمان تاخیر در مرخصی، تعداد دفعات عمل جراحی، فراوانی تشخیص، شدت بیماری، درجه تختخواب و نوع بیمه با مدت اقامت در بیمارستان ارتباط معناداری داشت. درک دقیق از عوامل مرتبط با مدت اقامت در بیمارستان و پیشرفت‌های تدریجی در پردازش و نظارت، ممکن است مدیریت کارآمدتر مدت اقامت بیماران بستری در بیمارستان را فراهم کند.

یانگ و همکاران (۲۰۱۵) [۱۲] در پژوهش خود از روش الگوریتم درخت تصمیم برای تحلیل داده‌ها در زمینه سرطان ریه استفاده کرده‌اند. در این پژوهش، از تکنیک‌های داده کاوی که برای تشخیص سرطان استفاده شده‌است، نمونه‌ای از بافت ریه بیمار برداشته می‌شود تا مورد تست‌های پاتولوژی قرار بگیرد. این مسئله اندازه و یا گستردگی تومور اصلی و همچنین اینکه سرطان توسعه یافته یا خیر را توصیف می‌کند. در این رویه، جراحی بافت برداری ضروری است اما ممکن است که وضعیت سلامتی بیمار را در معرض خطر قرار دهد. لذا این مطالعه متمرکز بر گرفتن اطلاعات بالینی‌ایی بوده که می‌توان بدون جراحی، برای جای گذاری گزارش پاتولوژی فراهم نمود. تکنیک‌های داده کاوی برای یافتن ارتباط مابین اطلاعات بالینی و گزارش پاتولوژی به‌منظور پشتیبانی از تشخیص مرحله پاتولوژیک سرطان ریه استفاده شده‌است.

گزارشات پاتولوژی پیش‌پردازش شده‌بود تا بدین‌منظور داده‌های تکراری طبقه‌بندی‌شده توسط روش درخت تصمیم حذف گردد. کاوش انجمنی داده، به‌عنوان ابزاری برای استخراج دانش از ارتباط مابین گزارشات پاتولوژی و اطلاعات بالینی بکار گرفته شده‌بود. برای این هدف، الگوریتم اperiوری به‌منظور استخراج قوانین انجمنی استفاده شده بود و اهمیت هر قانون تولیدشده، با استفاده از پشتیبانی، اعتماد و پیشرفت، آزمایش شده‌بود. هدف الگوریتم اperiوری، یافتن وابستگی‌ها بین مجموعه‌ای مختلف از داده است.

مریم طائفی و همکاران (۲۰۱۷) [۱۳] نیز از روش الگوریتم درخت تصمیم در بررسی بیماری عروق کرونر قلب استفاده کردند. بیماری عروق کرونر قلب یکی از مشکلات مهم بهداشت عمومی در سطح جهان است. الگوریتم‌هایی که شامل ارزیابی بیومارکرهای بالینی (یک نشانگر زیستی در حوزه پزشکی شاخصی قابل‌اندازه‌گیری برای تشخیص وجود یک بیماری است که گاهی شدت بیماری را نیز نشان می‌دهد) همراه با چندین فاکتور خطر شایع ثابت‌شده هستند، می‌توانند به پزشکان در پیش‌بینی بیماری عروق کرونر قلب و حمایت از تصمیم‌گیری‌های بالینی مربوط به مداخلات کمک کنند. درخت تصمیم یک مدل داده‌کاوی برای استخراج دانش پنهان از دیتابیس‌های بزرگ است. آنها برآنند تا با استفاده از الگوریتم درخت تصمیم، یک مدل پیش‌بینی‌کننده برای بیماری عروق کرونر قلب ایجاد کنند. این پژوهش تنها مطالعه الگوریتم‌های داده‌کاوی است که پروتئین واکنشی C فوق‌حساس سرمی را به‌عنوان یک ویژگی برای فاکتورهای خطر نتایج آنژیوگرافی در نظر می‌گیرد. آنها دریافتند میزان بالای پروتئین واکنشی C فوق‌حساس همراه با سطح قندخون در حالت ناشتا، جنسیت و سن، بیش از ۹۵ درصد از عوامل تعیین‌کننده حضور بیماری عروق کرونر قلب را تشکیل می‌دادند. این مدل می‌تواند فاکتورهای خطر مرتبط با بیماری عروق کرونر قلب را با حساسیت، صحت و دقت به‌ترتیب ۹۶ درصد، ۸۷ درصد، ۹۴ درصد شناسایی کند.

خواججه‌علی و همکاران (۲۰۱۷) [۱۴] نیز از روش الگوریتم درخت تصمیم در بررسی طول مدت بستری بیماران مبتلا به ذات‌الریه در بیمارستان‌ها استفاده کردند. ذات‌الریه عفونت شایع دستگاه تنفسی است که ریه‌ها را مبتلا می‌کند. متوسط طول مدت اقامت در بیمارستان یکی از ساده‌ترین و مهمترین شاخصه فعالیت بیمارستانی است که برای اهداف مختلف در نظر گرفته می‌شود. هدف از این پژوهش، بررسی عوامل مهم و موثر متوسط طول مدت اقامت بیماران مبتلا به ذات‌الریه در بیمارستان است. مدل پیشنهادی در این پژوهش نشان می‌دهد که چگونه سناریوهای مختلف بر معیار مدل بهره‌وری تاثیر می‌گذارند... لازم به ذکر است برخی از روش‌های بکار گرفته‌شده در این پژوهش، با تکنیک Bayesian boosting می‌تواند به نتایج بهتری دست یابد و عوامل موثر بر متوسط طول مدت اقامت را می‌توان با دقت ۹۵.۱۷ درصد شناسایی کرد. علاوه‌براین، ۷۴ درصد افراد مسن نسبت به بیماری ذات‌الریه آسیب‌پذیر بودند. همچنین مشخص شد مصرف مروپم در مقایسه با سایر آنتی‌بیوتیک‌های دیگر که برای درمان ذات‌الریه اکثر گروه‌های سنی مورد استفاده قرار می‌گیرند، باعث کاهش متوسط طول مدت اقامت بیماران می‌گردد. یکی از مسائل بهینه‌سازی، می‌تواند خوشه‌بندی اسناد باشد که با استفاده از الگوریتم‌های هوش دسته‌جمعی، می‌توان به خوشه‌بندی دقیق اسناد رسید [۱۵].

پایگاه داده‌ها در حوزه سلامت، حاوی میزان وسیعی از داده‌های بالینی است که کشف ارتباطات و الگوها در آن می‌تواند به دانش جدید پزشکی بیانجامد. امروزه با توجه به ظهور نظام‌های اطلاعات یکپارچه و رشد فناوری اطلاعات،

این مهم بیش از پیش نمایان شده است. داده کاوی از جمله پیشرفت‌های فناوری در راستای مدیریت داده‌ها است و استفاده گسترده از سیستم‌های اطلاعات و پایگاه‌های داده، ادغام آن را با شیوه‌های سنتی به یک الزام تبدیل کرده است. داده کاوی شاخه بسیار مهمی در فهم عمیق‌تر داده‌های پزشکی است که درصدد حل مسائل در تشخیص و درمان بیماری‌ها می‌باشد. با استفاده از داده کاوی، می‌توان بیماری‌های مزمن و خطرناک را شناسایی کرد و نتیجه این کار مدیریت بهتر بر روی بیماران این بیماری است که می‌توان به این بیماری‌ها اولویت بیشتری اختصاص داده و در اورژانس‌ها و موقعیت‌هایی که زمان و منابع موردنیاز در برخورد با تمام بیماران محدود می‌باشد، این بیماران را در اولویت اول قرار داد. حتی می‌توان تعداد موارد بستری در بیمارستان‌ها را کاهش داد و از پذیرش‌های بی‌جهت جلوگیری کرد. همچنین می‌توان عوارض دارویی و تضاد دارویی برخی از درمان‌ها را شناسایی کرد و برای درمان دیگر بیماران از نتیجه این الگوها استفاده نمود.

حوزه پزشکی و سلامت، از بخش‌های مهم در جوامع صنعتی است. استخراج داده‌هایی از میان حجم انبوه داده‌های مرتبط با سوابق بیماری و پرونده‌های پزشکی افراد با استفاده از فرآیند داده کاوی، می‌تواند منجر به شناسایی قوانین حاکم بر ایجاد و رشد بیماری‌ها شده و اطلاعات ارزشمندی را به منظور شناسایی علل رخداد بیماری‌ها، تشخیص، پیش‌بینی و درمان بیماری‌ها با توجه به عوامل محیطی حاکم در اختیار متخصصان و دست‌اندرکاران حوزه سلامت قرار دهد. نتیجه این مسئله به معنای افزایش عمر و ایجاد آرامش برای افراد جامعه است. برای دسته‌بندی اطلاعات می‌توان از روش‌های یادگیری ماشین استفاده کرد. با این روش و چند ویژگی مختلف دیگر، می‌توان یک کلاس‌بندی به منظور تجزیه و تحلیل داده‌ها ایجاد نمود [۱۶].

دانش داده کاوی از جمله روش‌های موثری است که می‌تواند از اطلاعات و داده‌های جاری در نظام سلامت به عنوان ابزاری قدرتمند برای پیشگیری، درمان، کاهش هزینه‌ها، تصمیم‌گیری مناسب و ارائه خدمات مناسب به مشتریان سلامت، استفاده کند. چراکه در محیط رقابتی امروز، سازمان‌هایی موفق هستند که برای بهبود کیفیت از بهترین ابزارها بهره ببرند [۱۷]. در موضوعات آموزشی و بهبود فرآیندها این دانش کمک زیادی نموده است [۱۸] لذا در این پژوهش راهکارهای افزایش رضایتمندی بیماران بستری شده با استفاده از تکنیک‌های داده کاوی بررسی می‌شود.

با توجه به موارد فوق و مرور مقالات ارائه شده در این زمینه، می‌توان بیان نمود اگرچه مطالعات پیشین بر روی بررسی طول اقامت بیماران تمرکز داشته‌اند، این پژوهش برای اولین بار از روش‌های قوانین انجمنی باعث کشف قوانین مرتبط با طول اقامت بیماران شده است.

در این خصوص می‌توان سئوالات اصلی تحقیق را به صورت زیر بیان نمود:

آیا بین متغیرهای زیر و متوسط طول مدت اقامت رابطه معناداری وجود دارد؟

0 بستری شدن در بخش اورژانس

0 تزریق خون

0 عمل جراحی

0 دستور پزشک

(۳) روش تحقیق

تمام داده‌های جمع‌آوری شده از بیمارستان امیرالمومنین مراغه شامل ۲۶۹۰۷ بیمار بود. متغیرها به ترتیب عبارتند از: جنسیت، وضعیت تاهل، بخش بستری، تخصص پزشک، بیمه، تزریق خون، عمل جراحی، نوع ترخیص، سن و متوسط طول مدت اقامت که ویژگی هر کدام از متغیرها در جدول ۱ آورده شده است.

جدول ۱. متغیرها

نوع متغیر	عنوان
گسسته	جنسیت
گسسته	وضعیت تاهل
گسسته	بخش بستری
گسسته	تخصص پزشک معالج
گسسته	بیمه
گسسته	تزریق خون
گسسته	عمل جراحی
گسسته	نحوه ترخیص
عددی	سن

جهت تحلیل داده‌ها، گام اول خوشه‌بندی و گروه‌بندی اطلاعات بیماران است.

(۳-۱) خوشه‌بندی

خوشه‌بندی، سازماندهی مجموعه‌ای از الگوها براساس شباهت در خوشه‌ها است. به نحوی که الگوهای داخل یک خوشه شبیه به هم بوده و دارای بیشترین تفاوت با الگوهای خوشه‌های دیگر باشند. این وظیفه اصلی داده‌کاوی اکتشافی است و یک روش معمول برای تجزیه و تحلیل داده‌های آماری است که در بسیاری از زمینه‌ها از جمله یادگیری ماشین، تشخیص الگو، تجزیه و تحلیل تصویر، بازیابی اطلاعات، بیوانفورماتیک، فشرده‌سازی داده‌ها و گرافیک کامپیوتری استفاده می‌شود. تجزیه و تحلیل خوشه‌ای خود یک الگوریتم خاص نیست بلکه روند کلی است و می‌تواند توسط الگوریتم‌های مختلفی به دست آید که در درک آنچه یک خوشه را تشکیل می‌دهند و نحوه کارآمدی آنها را پیدا می‌کند. اصطلاحات خوشه‌ها شامل گروه‌هایی با فاصله‌های کم بین اعضای خوشه، مناطق متراکم فضای داده، فواصل و یا توزیع‌های آماری خاص است. بنابراین خوشه‌بندی می‌تواند به عنوان یک مسئله بهینه‌سازی چندهدفه صورت گیرد. الگوریتم خوشه‌بندی مناسب و تنظیمات پارامتر (از جمله پارامترهایی مانند تابع فاصله مورد استفاده، آستانه تراکم یا تعداد خوشه مورد انتظار) بستگی به تنظیم مجموعه داده‌ها توسط فرد و استفاده خاص فرد از نتایج دارد. تجزیه و تحلیل خوشه‌ای یک روش اتوماتیک نیست بلکه یک فرآیند

تکراری از کشف دانش یا بهینه‌سازی چندهدفه تعاملی است که شامل آزمایش و شکست است. اغلب لازم است که داده‌های پیش‌پردازش شده و پارامترهای مدل اصلاح شوند تا نتیجه حاصل، همان نتیجه دلخواه باشد. چهار گروه اصلی برای الگوریتم‌های خوشه‌بندی وجود دارد. الگوریتم‌های خوشه‌بندی تفکیکی ۱، الگوریتم‌های خوشه‌بندی سلسله‌مراتبی ۲، الگوریتم‌های خوشه‌بندی بر مبنای چگالی ۳ و الگوریتم‌های خوشه‌بندی بر مبنای مدل ۴. در این پژوهش جهت مدل‌سازی از الگوریتم خوشه‌بندی کامینز ۵ که از گروه الگوریتم‌های خوشه‌بندی تفکیکی است، استفاده می‌شود و درجه پیچیدگی محاسباتی آن برابر با $O(ndk+1)$ است، به شرطی که n تعداد اشیای، d بعد ویژگی‌ها و k تعداد خوشه‌ها باشد.

۳-۲) استخراج دانش قوانین انجمنی

فرآیند کشف قوانین وابستگی، یکی از رویکردهای مهم در علم نوین داده کاوی برای یافتن قواعد و الگوها در پایگاه داده است. در گام استخراج دانش، هدف تولید قواعد انجمنی از مجموعه داده است تا روابط بین ویژگی‌ها استخراج شوند. فرضیات ابتدایی عبارتند از: مجموعه $I = I_1, I_2, I_3, \dots$ به عنوان مجموعه اقلام در نظر گرفته می‌شوند. مجموعه D هم به عنوان مجموعه داده‌ها یعنی تراکنش‌های پایگاه داده‌ها در نظر گرفته می‌شوند بطوریکه هر تراکنش شامل مجموعه‌ای از اقلام باشد یعنی هر تراکنش T زیرمجموعه‌ای از I است. هر تراکنش شناسه‌ای به نام TID دارد. اگر A یک مجموعه از اقلام باشد، گفته می‌شود تراکنش T شامل A است اگر و فقط اگر $A \subseteq T$ باشد. یک قانون انجمنی گزاره‌ای است به صورت $A \Rightarrow B$ بطوریکه $A \subseteq I$ ، $B \subseteq I$ و $A \cap B = \emptyset$.

معیار قبولیت قوانین انجمنی دو پارامتر مهم یعنی $\text{support} = 6$ و $\text{Confidence} = 2$ در نظر گرفته می‌شود. طبق عبارت (۱) قانون $A \Rightarrow B$ در مجموعه تراکنش‌های D دارای پشتیبانی برابر با s است. اگر s درصد از تراکنش‌های D شامل $A \cup B$ باشد و این قانون دارای قابلیت اطمینان برابر c است. c درصد از تراکنش‌هایی که شامل A هستند شامل B نیز هستند.

$$\text{Support}(A \Rightarrow B) = P(A \cup B)$$

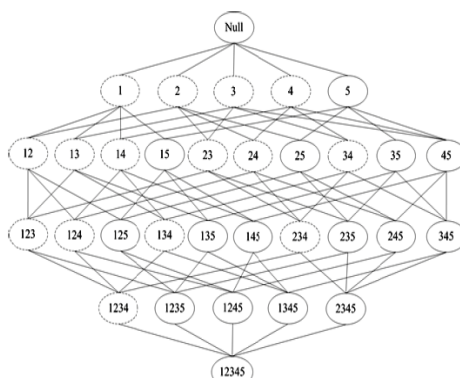
$$\text{Confidence}(A \Rightarrow B) = P(B | A)$$

(۱)

قوانینی که حد پایین Support یا min-sup و حد پایین Confidence یا min-conf را دارا باشند، قوانین انجمنی قوی نامیده می‌شوند و هدف تمام الگوریتم‌ها چنین قوانینی است. در بیشتر الگوریتم‌ها از جمله اپریوری، پیدا کردن قوانین انجمنی، فرآیندی است دو مرحله‌ای. در مرحله اول: مجموعه اقلام مهم یعنی آنهایی که دارای Support بیشتر از min-sup هستند، شناسایی می‌شوند و در مرحله دوم: از این مجموعه اقلام برای تولید قوانین انجمنی قوی استفاده می‌شود. نکته جالب توجه این است که کارآیی کلی الگوریتم‌ها، کشف قوانین انجمنی وابسته به مرحله اول فرآیند است و پس از پیدا شدن اقلام داده‌ای مهم، قوانین انجمنی به سادگی از روی آنها قابل استخراج خواهد بود.

۳-۳) الگوریتم اپریوری^۱

الگوریتم اپریوری، یک الگوریتم قدرتمند برای کاوش مجموعه آیت‌های پرتکرار برای کشف قوانین انجمنی می‌باشد. همچنین روشی قابل‌اعمال روی رکوردهای پایگاه داده به‌ویژه پایگاه داده تراکنشی یا رکوردهای حاوی تعداد مشخصی فیلد یا آیت است. اپریوری یکی از الگوریتم‌های دارای رویکرد «پایین به بالا» است که به تدریج رکوردهای پیچیده را با یکدیگر مقایسه می‌کند. این الگوریتم یکی از روش‌های کارآمد برای حل مسائل پیچیده کنونی موجود در داده‌کاوی و یادگیری ماشین است. اساساً الگوریتم اپریوری، بخش‌هایی از یک پایگاه داده بزرگتر را دریافت کرده و به آنها امتیازدهی کرده و یا آن بخش‌ها را با دیگر مجموعه‌ها به شیوه مرتب‌شده‌ای مقایسه می‌کند. از نتایج خروجی، برای تولید مجموعه‌هایی استفاده می‌شود که مکرراً در پایگاه داده اصلی به‌وقوع پیوسته‌اند. الگوریتم اپریوری یک الگوریتم جستجوی سطحی است که با پایان کاوش در مرحله k ام به مرحله بعدی یعنی $k+1$ می‌رود. این عمل تا محقق شدن شرط یا شروط پایانی تکرار می‌شود. در مرحله k ام مجموعه اقلام k تایی تولید خواهند شد. پس از محاسبه مقدار Support برای هر کدام و مقایسه آن با مقدار minsup الگوی‌های مکرر k تایی شناسایی می‌شود. در مرحله بعدی الگوریتم با کمک الگوهای مکرر k تایی، مجموعه اقلام $k+1$ تایی کاندید که می‌توانند مکرر باشند را تولید می‌کند. به‌همین ترتیب با توجه به مقدار minsup برخی حذف شده و مجموعه اقلام مکرر $k+1$ تایی تشکیل خواهند شد. این عمل تا یافتن آخرین مجموعه قلم مکرر ادامه پیدا می‌کند. این الگوریتم در حین اجرا از قاعده‌ای موسوم به قاعده اپریوری استفاده می‌کند که بدین صورت بیان می‌شود: «اگر یک الگوی مکرر داشته باشیم، کلیه زیرمجموعه‌های آن نیز مکرر هستند.» به عبارت دیگر، اگر مجموعه اقلام I مکرر نباشد، هر مجموعه‌ای که شامل I است نیز نمی‌تواند مکرر باشد. با کمک قاعده اپریوری فضای جستجو کاهش می‌یابد. شکل زیر کل فضای جستجو را برای مجموعه اقلامی که از $\{1, 2, 3, 4, 5\}$ استفاده کرده‌است را نشان می‌دهد. برای سادگی الگوها به شکل یک عدد نمایش داده شده‌است. برای مثال الگوی $\{1, 2, 3\}$ به صورت ۱۲۳ نشان داده شده‌است. در شکل الگوهای مکرر با دایره‌های خط‌چین مشخص شده‌است. با نگاه به الگوی مکرر $\{1, 2, 3, 4\}$ که در شکل به صورت ۱۲۳۴ است، همه زیرمجموعه‌های آن نیز مکرر هستند یا اینکه تمام مجموعه اقلامی که شامل $\{5\}$ هستند، نمی‌توانند مکرر باشند چون $\{5\}$ مکرر نیست. بدین ترتیب استفاده از این استراتژی باعث کاهش فضای جستجو در تولید اقلام مکرر می‌شود.



شکل ۱. فضای جستجو

۴) یافته‌ها

۴-۱) تجزیه و تحلیل خوشه‌بندی‌های دو تا پنج

جهت تجزیه و تحلیل از نرم‌افزار وکا استفاده شده است. نرم‌افزار وکا (Weka) یک ابزار متن‌باز برای یادگیری ماشین و داده کاوی است که الگوریتم‌های مختلفی مانند خوشه‌بندی و قواعد انجمنی را پیاده‌سازی می‌کند.

در تجزیه و تحلیل خوشه‌بندی، گروه‌بندی مجموعه‌ای از اشیاء انجام می‌شود. این کار به این صورت است که اشیاء موجود در یک گروه (به نام خوشه) در مقایسه با دیگر دسته‌ها (خوشه‌ها) مشابه‌تر هستند. در این پژوهش، خوشه‌بندی با استفاده از روش K-Means با تنظیم تعداد خوشه دو تا پنج انتخاب شده است و هر کدام از خوشه‌ها، تک‌تک تحلیل می‌شوند تا دید بازتری برای انتخاب بهترین تعداد خوشه به بیمارستان ارائه شود. همچنین جهت پیاده‌سازی قواعد انجمنی در نرم‌افزار وکا از تکنیک Apriori استفاده شده است. تعداد خوشه‌ها برای K-Means، حداقل اطمینان برای Apriori پارامترهای مهم برای ارزیابی به‌شمار می‌رود.

مطابق شکل ۲، برای دو خوشه ۴۱ درصد بیماران (۱۰۵۵) بیمار در خوشه صفر و ۵۹ درصد بیماران (۱۵۸۵۲) بیمار در خوشه یک قرار گرفته است. اکثر تعداد هر دو خوشه مرد هستند. میانگین سنی در خوشه صفر تقریباً ۳۴ و در خوشه یک تقریباً ۴۴ می‌باشد. اکثر بیمارانی که در خوشه صفر قرار گرفته‌اند، مجردند و اکثر بیماران متأهل در خوشه یک دسته‌بندی شده‌اند. متوسط طول مدت اقامت در خوشه صفر ۱/۲۵۹ و در خوشه یک ۱/۵۸۱ روز می‌باشد.

Final cluster centroids:			
Attribute	Full Data (26907.0)	Cluster#	
		0 (11055.0)	1 (15852.0)
number	13454	14870.4595	12466.1776
SexType	man	man	man
age	40.24	34.251	44.4167
MaritalType	Married	Single	Married
secname	emergency department	emergency department	emergency department
Speciality Name	Emergency Medicine	Emergency Medicine	Emergency Medicine
binname	ok	ok	ok
khon	NOK	NOK	NOK
amal	NOK	NOK	NOK
removecause	Doctors order	Doctors order	Doctors order
LOS	1.4496	1.2598	1.5819
Time taken to build model (full training data) : 0.36 seconds			
=== Model and evaluation on training set ===			
Clustered Instances			
0	11055 (41%)		
1	15852 (59%)		

شکل ۲. خروجی نرم‌افزار وکا برای دو خوشه

خروجی روش کامینز که مطابق شکل ۳ با استفاده از نرم‌افزار وکا برای تعداد خوشه سه به‌دست آمده است، نشان می‌دهد که طبق این خوشه‌بندی ۲۶ درصد بیماران (۶۸۶۴) بیمار در خوشه صفر و ۴۹ درصد بیماران (۱۳۲۶۱) بیمار در خوشه یک و ۲۵ درصد بیماران (۶۷۸۲) بیمار در خوشه دو قرار گرفته‌اند. اکثر بیماران در خوشه یک دسته‌بندی شده‌اند. در خوشه صفر تعداد بیشتری از بیماران زن هستند و در خوشه یک و دو اکثر بیماران مرد هستند. میانگین سنی در خوشه

صفر تقریباً ۳۶ و در خوشه یک تقریباً ۴۴ و در خوشه دو ۳۸ می‌باشد. اکثر بیمارانی که در خوشه صفر قرار گرفته‌اند مجردند و اکثر بیماران متاهل در خوشه‌های یک و دو دسته‌بندی شده‌اند. اکثر بیمارانی که عمل شده‌اند، در خوشه دو دسته‌بندی شده‌اند. متوسط طول مدت اقامت در خوشه دو بالاترین نسبت به دو خوشه بعدی ۲/۴۲ روز و خوشه‌های صفر و یک در حدود ۱/۱۲ روز می‌باشد.

Final cluster centroids:				
Attribute	Full Data (26907.0)	Cluster#		
		0 (6864.0)	1 (13261.0)	2 (6782.0)
number	13454	20057.4024	14875.7393	3990.7976
SexType	man	woman	man	man
age	40.24	36.1502	43.6147	37.7807
MaritalType	Married	Single	Married	Married
secname	emergency department	emergency department	emergency department	Mens surgery department
Specialty Name	Emergency Medicine	Emergency Medicine	Emergency Medicine	Orthopedics
binname	ok	ok	ok	ok
khon	NOK	NOK	NOK	NOK
amal	NOK	NOK	NOK	NOK
removecause	Doctors order	Doctors order	Doctors order	Doctors order
LOS	1.4496	1.1225	1.1109	2.4272

Time taken to build model (full training data) : 0.58 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	6864 (26%)
1	13261 (49%)
2	6782 (25%)

شکل ۳. خروجی نرم‌افزار وکا برای سه خوشه

خروجی روش کامینز که مطابق شکل ۴ با استفاده از نرم‌افزار وکا برای تعداد خوشه چهار به‌دست آمده‌است، قابل‌ملاحظه است که طبق این خوشه‌بندی ۲۵ درصد بیماران (۶۸۳۳ بیمار) در خوشه صفر، ۴۲ درصد بیماران (۱۱۱۶۷ بیمار) در خوشه یک، ۲۵ درصد بیماران (۶۷۳۰ بیمار) در خوشه دو و ۸ درصد بیماران (۲۱۷۷) در خوشه سه قرار گرفته‌اند. اکثر بیماران در خوشه یک دسته‌بندی شده‌اند. در خوشه صفر تعداد بیشتری از بیماران زن هستند و در خوشه یک، دو و سه اکثر بیماران مرد هستند. میانگین سنی در خوشه صفر تقریباً ۳۶، در خوشه یک تقریباً ۴۵، در خوشه دو تقریباً ۳۸ و در خوشه سه ۳۴ می‌باشد. اکثر بیمارانی که در خوشه صفر قرار گرفته‌اند مجردند و اکثر بیماران متاهل در خوشه‌های یک، دو و سه دسته‌بندی شده‌اند. بیمارانی که در خوشه سه قرار دارند اکثراً فاقد بیمه هستند. اکثر بیمارانی که عمل شده‌اند، در خوشه دو دسته‌بندی شده‌اند. متوسط طول مدت اقامت در خوشه دو بالاترین نسبت به سه خوشه بعدی یعنی ۲/۴۳ روز و خوشه سه با کمترین مدت اقامت یعنی ۰/۲۸ می‌باشد.

Final cluster centroids:					
Attribute	Full Data (26907.0)	Cluster# 0 (6833.0)	1 (11167.0)	2 (6730.0)	3 (2177.0)
number	13454	19933.7021	14364.575	3985.5048	17715.9343
SexType	man	woman	man	man	man
age	40.24	36.1043	45.4023	37.742	33.9245
MaritalType	Married	Single	Married	Married	Married
secname	emergency department	emergency department	emergency department Mens surgery department	emergency department	emergency department
Speciality Name	Emergency Medicine	Emergency Medicine	Emergency Medicine	Orthopedics	Emergency Medicine
binname	ok	ok	ok	ok	ok
khon	NOK	NOK	NOK	NOK	NOK
amal	NOK	NOK	NOK	OK	NOK
removecause	Doctors order	Doctors order	Doctors order	Doctors order	Doctors order
LOS	1.4496	1.1442	1.271	2.4349	0.274
Time taken to build model (full training data) : 0.31 seconds					
=== Model and evaluation on training set ===					
Clustered Instances					
0	6833 (25%)				
1	11167 (42%)				
2	6730 (25%)				
3	2177 (8%)				

شکل ۴. خروجی نرم‌افزار وکا برای چهار خوشه

خروجی روش کامینز که مطابق شکل ۵ با استفاده از نرم‌افزار وکا برای تعداد خوشه پنج به‌دست آمده‌است، قابل‌ملاحظه است که طبق این خوشه‌بندی ۱۵ درصد بیماران (۴۱۰۸ بیمار) در خوشه صفر، ۳۱ درصد بیماران (۸۴۶۵ بیمار) در خوشه یک، ۲۵ درصد بیماران (۶۶۴۱ بیمار) در خوشه دو، ۶ درصد بیماران (۱۶۱۸) در خوشه سه و ۲۳ درصد بیماران (۶۰۷۵ بیمار) در خوشه چهار قرار گرفته‌اند. اکثر بیماران در خوشه یک دسته‌بندی شده‌اند. در خوشه چهار تعداد بیشتری از بیماران زن هستند و در خوشه یک، دو، سه و چهار اکثر بیماران مرد هستند. میانگین سنی در خوشه صفر تقریباً ۳۷، در خوشه یک تقریباً ۴۷، در خوشه دو تقریباً ۳۸، در خوشه سه ۳۴ و در خوشه چهار ۳۷ می‌باشد. اکثر بیمارانی که در خوشه صفر و چهار قرار گرفته‌اند، مجردند و اکثر بیماران متاهل در خوشه‌های یک، دو و سه دسته‌بندی شده‌اند. بیمارانی که در خوشه سه قرار دارند، اکثراً فاقد بیمه هستند. اکثر بیمارانی که عمل شده‌اند، در خوشه دو دسته‌بندی شده‌اند. متوسط طول مدت اقامت در خوشه دو بالاترین نسبت به سه خوشه بعدی یعنی ۲/۴۳ روز و خوشه سه با کمترین مدت اقامت یعنی ۰/۲۸ می‌باشد.

Final cluster centroids:							
Attribute	Full Data (26907.0)	Cluster# 0 (4108.0)	1 (8465.0)	2 (6641.0)	3 (1618.0)	4 (6075.0)	
number	13454	13205.4276	15035.2967	4000.4379	19401.9438	20114.4258	
SexType	man	man	man	man	man	woman	
age	40.24	36.7802	46.9044	37.941	34.3523	37.3748	
MaritalType	Married	Single	Married	Married	Married	Single	
secname	emergency department	emergency department	emergency department Mens surgery department	emergency department	emergency department	emergency department	
Speciality Name	Emergency Medicine	Emergency Medicine	Emergency Medicine	Orthopedics	Emergency Medicine	Emergency Medicine	
binname	ok	ok	ok	ok	ok	ok	
khon	NOK	NOK	NOK	NOK	NOK	NOK	
amal	NOK	NOK	NOK	OK	NOK	NOK	
removecause	Doctors order	Doctors order	Doctors order	Doctors order	Doctors order	Doctors order	
LOS	1.4496	0.8578	1.3388	2.4319	0.2738	1.2435	
Time taken to build model (full training data) : 0.27 seconds							
=== Model and evaluation on training set ===							
Clustered Instances							
0	4108 (15%)						
1	8465 (31%)						
2	6641 (25%)						
3	1618 (6%)						
4	6075 (23%)						

شکل ۵. خروجی نرم‌افزار وکا برای پنج خوشه

نتایج به‌دست آمده از خوشه‌های مختلف را می‌توان در قالب جدول (۲) خلاصه نمود:

جدول ۲. طول مدت اقامت در خوشه‌بندی‌های متفاوت

تعداد خوشه	کمترین طول اقامت (روز)	بیشترین طول اقامت (روز)
۲	خوشه صفر ۱/۲۵۹	در خوشه یک ۱/۵۸۱
۳	خوشه‌های صفر و یک در حدود ۱/۱۲	خوشه دو ۲/۴۲
۴	خوشه سه ۰/۲۸	خوشه دو ۲/۴۳
۵	خوشه سه ۰/۲۸	خوشه دو ۲/۴۳

(۵) نتیجه‌گیری و پیشنهادها

قواعد انجمنی، یکی از تکنیک‌های اصلی داده کاوی است و تقریباً مهمترین شکل کشف و استخراج الگوها در سیستم‌های یادگیری می‌باشد. قوانین انجمنی، ارتباطات جذاب بین مجموعه بزرگی از داده‌ها را کشف می‌کند که این ارتباطات می‌تواند به تصمیم‌گیرندگان کمک کند. درواقع قوانین انجمنی شرایطی را نشان می‌دهند که در یک مجموعه داده، مکرراً باهم اتفاق می‌افتد. در این پژوهش، از الگوریتم اپریوری برای استخراج قواعد انجمنی استفاده شد. مطابق با شکل ۶، این قوانین با $\text{support} = 0.65$ و $\text{Confidence} = 0.80$ اجرا شده‌است. موارد زیر ازجمله قوانین مهم استخراج‌شده الگوریتم اپریوری توسط نرم‌افزار و کا می‌باشد:

۱. با اطمینان ۱۰۰ درصد؛ اگر بیمار در بخش اورژانس بستری شده باشد، تحت عمل جراحی قرار نخواهد گرفت.
۲. با اطمینان ۱۰۰ درصد؛ اگر بیمار در بخش اورژانس بستری شده باشد و تحت تزریق خون قرار نگرفته باشد، تحت عمل جراحی قرار نخواهد گرفت.
۳. با اطمینان ۱۰۰ درصد؛ اگر بیمار در بخش اورژانس بستری شده باشد، تحت تزریق خون قرار نخواهد گرفت.
۴. با اطمینان ۱۰۰ درصد؛ اگر بیمار در بخش اورژانس بستری شده باشد و تحت عمل جراحی قرار نگیرد، آنگاه تحت تزریق خون قرار نخواهد گرفت.
۵. با اطمینان ۱۰۰ درصد؛ اگر بیمار در بخش اورژانس بستری شده باشد، تحت عمل جراحی و تزریق خون قرار نخواهد گرفت.
۶. با احتمال ۱۰۰ درصد؛ بیماری که تحت عمل جراحی قرار نگرفته باشد و با دستور پزشک از بیمارستان مرخص شده باشد، تحت تزریق خون قرار نخواهد گرفت.
۷. با احتمال ۱۰۰ درصد؛ بیماری که با دستور پزشک از بیمارستان مرخص شود، نیاز به تزریق خون نخواهد داشت.
۸. با احتمال ۱۰۰ درصد؛ بیماری که تحت عمل جراحی نبوده‌است، تحت تزریق خون نیز نخواهد بود.
۹. با احتمال ۹۹ درصد؛ بیماری که بیمه دارد و تحت تزریق خون قرار نگرفته‌است، با دستور پزشک از بیمارستان مرخص خواهد شد.
۱۰. با احتمال ۹۹ درصد؛ بیماری که تحت تزریق خون قرار نگرفته‌است، با دستور پزشک از بیمارستان مرخص خواهد شد.

۱۱. با احتمال ۹۹ درصد؛ بیماری که تحت تزریق خون و نیز عمل جراحی قرار نگرفته‌است، با دستور پزشک از بیمارستان مرخص خواهد شد.

۱۲. با احتمال ۹۸ درصد؛ بیمارانی که تحت عمل جراحی قرار گرفته‌اند، با دستور پزشک از بیمارستان مرخص خواهند شد.

۱۳. با احتمال ۹۸ درصد؛ بیماری که تحت عمل جراحی قرار نگرفته‌است، تحت تزریق خون قرار نخواهد گرفت و نیز با دستور پزشک از بیمارستان مرخص خواهد شد.

۱۴. با احتمال ۹۸ درصد؛ بیماری که بیمه داشته باشد، تحت تزریق خون قرار نخواهد گرفت و نیز با دستور پزشک از بیمارستان مرخص خواهد شد.

۱۵. با احتمال ۸۸ درصد؛ بیماری که با دستور پزشک از بیمارستان مرخص شده‌است، بیمه بوده‌است.

۱۶. با احتمال ۸۸ درصد؛ بیماری که تحت تزریق خون قرار نگرفته‌است و همچنین با دستور پزشک از بیمارستان ترخیص شده‌است، بیمه بوده‌است.

```
Apriori
=====

Minimum support: 0.65 (17490 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 7

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 8

Size of set of large itemsets L(3): 4

Best rules found:

1. Inpatient department=emergency department 17514 ==> Surgical=No 17513 <conf:(1)> lift:(1.31) lev:(0.15) [4132] conv:(2066.96)
2. Inpatient department=emergency department 17510 ==> Surgical=No 17509 <conf:(1)> lift:(1.31) lev:(0.15) [4131] conv:(2066.49)
3. Inpatient department=emergency department 17514 ==> Blood injection=No 17510 <conf:(1)> lift:(1) lev:(0) [63] conv:(13.41)
4. Inpatient department=emergency department 17513 ==> Blood injection=No 17509 <conf:(1)> lift:(1) lev:(0) [63] conv:(13.41)
5. Inpatient department=emergency department 17514 ==> Blood injection=No Surgical=No 17509 <conf:(1)> lift:(1.31) lev:(0.16) [4171] conv:(696.04)
6. Surgical=No Clearance=Doctors order 20234 ==> Blood injection=No 20189 <conf:(1)> lift:(1) lev:(0) [32] conv:(1.68)
7. Clearance=Doctors order 26475 ==> Blood injection=No 26410 <conf:(1)> lift:(1) lev:(0) [36] conv:(1.54)
8. binnam=Yes Clearance=Doctors order 23289 ==> Blood injection=No 23224 <conf:(1)> lift:(1) lev:(0) [24] conv:(1.35)
9. Surgical=No 20556 ==> Blood injection=No 20491 <conf:(1)> lift:(1) lev:(0) [13] conv:(1.19)
10. binnam=Yes Surgical=No 17633 ==> Blood injection=No 17568 <conf:(1)> lift:(1) lev:(0) [2] conv:(1.02)
11. binnam=Yes 23667 ==> Blood injection=No 23564 <conf:(1)> lift:(1) lev:(-0) [-12] conv:(0.87)
12. binnam=Yes Blood injection=No 23564 ==> Clearance=Doctors order 23224 <conf:(0.99)> lift:(1) lev:(0) [38] conv:(1.11)
13. Blood injection=No 26804 ==> Clearance=Doctors order 26410 <conf:(0.99)> lift:(1) lev:(0) [36] conv:(1.09)
14. Blood injection=No Surgical=No 20491 ==> Clearance=Doctors order 20189 <conf:(0.99)> lift:(1) lev:(0) [26] conv:(1.09)
15. Surgical=No 20556 ==> Clearance=Doctors order 20234 <conf:(0.98)> lift:(1) lev:(0) [8] conv:(1.02)
16. binnam=Yes 23667 ==> Clearance=Doctors order 23289 <conf:(0.98)> lift:(1) lev:(0) [1] conv:(1)
17. Surgical=No 20556 ==> Blood injection=No Clearance=Doctors order 20189 <conf:(0.98)> lift:(1) lev:(0) [12] conv:(1.03)
18. binnam=Yes 23667 ==> Blood injection=No Clearance=Doctors order 23224 <conf:(0.98)> lift:(1) lev:(-0) [-5] conv:(0.98)
19. Clearance=Doctors order 26475 ==> binnam=Yes 23289 <conf:(0.88)> lift:(1) lev:(0) [1] conv:(1)
20. Blood injection=No Clearance=Doctors order 26410 ==> binnam=Yes 23224 <conf:(0.88)> lift:(1) lev:(-0) [-5] conv:(1)
```

شکل ۶. خروجی الگوریتم اِپریوری توسط نرم‌افزار وکا

۶) ارزیابی نتایج

در این تحقیق جهت ارزیابی خوشه‌بندی، از معیار نیم‌رخ ۱ استفاده شده‌است. این معیار هم به پیوست درون خوشه‌ها و هم به میزان تفکیک پذیری آنها بستگی دارد. مقدار نیم‌رخ برای هر نقطه، میزان تعلق آن را به خوشه‌اش در مقایسه با خوشه مجاور اندازه می‌گیرد.

فرض کنید نقطه‌ای مانند x_i در میان داده‌هایی که خوشه‌بندی شده‌است، وجود دارد و در طی مراحل خوشه‌بندی نیز k خوشه $(C_1, C_2, \dots, C_k, C_1, C_2, \dots, C_k)$ ایجاد شده‌است، میانگین فاصله یک نقطه از خوشه با نقاط دیگر آن خوشه اندازه‌گیری می‌گردد. در این تحقیق از بین حالات متعدد برای ۲ تا ۵ خوشه، ۵ خوشه معیار نیم‌رخ کمتری را داشته‌است. به منظور ارزیابی الگوریتم‌های کشف قوانین انجمنی، از آنجاییکه این الگوریتم‌ها پتانسیل این را دارد که الگوها و قوانین زیادی تولید نماید، جهت ارزیابی این قوانین به عواملی همچون شخص استفاده‌کننده از قوانین و نیز حوزه‌ای که مجموعه داده مورد بررسی به آن تعلق دارد، وابستگی زیادی دیده می‌شود و بدین ترتیب کار پیدا کردن قوانین جذاب، به آسانی میسر نیست. معیار Support دارای مقدار عددی بین صفر و یک می‌باشد و هرچه این میزان بیشتر باشد، نشان می‌دهد که این دو شیء بیشتر باهم در ارتباط هستند. کاربر می‌تواند با مشخص کردن یک آستانه برای این معیار، تنها قوانینی را به دست آورد که Support آنها بیشتر از مقدار آستانه باشد. بدین ترتیب می‌توان با کاهش فضای جستجو، زمان لازم جهت پیدا کردن قوانین انجمنی را کمینه کرد. البته باید به ضعف این روش نیز توجه داشت که ممکن است قوانین با ارزشی را بدین ترتیب از دست بدهیم. درواقع استفاده از این معیار به تنهایی کافی نیست. طبق عبارت (۲) معیار Confidence نیز مقدار عددی بین صفر و یک می‌باشد که هرچه این عدد بزرگتر باشد، بر کیفیت قانون، افزوده خواهد شد. استفاده از این معیار به همراه Support مکمل مناسبی برای ارزیابی قوانین انجمنی خواهد بود. مشکلی که همچنان وجود دارد این است که امکان دارد قانونی با Confidence بالا وجود داشته باشد ولی از نظر ما ارزشمند نباشد. از معیارهای دیگر قوانین انجمنی می‌توان به معیار Lift اشاره کرد که این معیار میزان استقلال میان دو طرف قانون را نشان می‌دهد و می‌تواند مقدار عددی بین صفر تا بی‌نهایت باشد. درواقع Lift میزان هم‌اتفاقی بین ویژگی‌ها را در نظر می‌گیرد. مقادیر نزدیک به عدد یک، معرف این است که طرفین قانون مستقل از یکدیگر می‌باشند لذا نشان‌دهنده قانون جذابی نمی‌باشند. چنانچه این معیار از عدد یک کمتر باشد، نشان‌دهنده این است که طرفین قانون با یکدیگر رابطه منفی دارند. درضمن این معیار نسبت به سمت چپ و راست قانون متقارن است. درواقع اگر سمت چپ و راست قانون را با یکدیگر جابجا کنیم، مقدار این معیار تغییری نمی‌کند. از آنجاییکه این معیار نمی‌تواند به تنهایی برای ارزیابی مورد استفاده قرار گیرد و حتماً باید در کنار معیارهای دیگر باشد، باید مقادیر آن بین بازه صفر و یک نرمال شود. ترکیب این معیار به همراه Support و Confidence جزء بهترین روش‌های کاوش قوانین انجمنی است. مشکل این معیار حساس بودن به تعداد نمونه‌های مجموعه داده به‌ویژه برای مجموعه تراکنش‌های کوچک می‌باشد. از این رو معیارهای دیگری برای جبران این نقص معرفی شده‌اند. معیار Conviction برخی ضعف‌های معیارهای Confidence و Lift را جبران می‌نماید. محدوده قابل تعریف برای این معیار در حوزه ۰/۵ تا بی‌نهایت قرار می‌گیرد که هرچه این مقدار بیشتر باشد، نشان‌دهنده این است که آن قانون جذاب‌تر است. بر خلاف Lift این معیار متقارن نمی‌باشد و مقدار این معیار برای دلالت‌های منطقی یعنی در جاییکه Confidence قانون یک می‌باشد برابر با بی‌نهایت است و چنانچه طرفین قانون مستقل ازهم باشند، مقدار این معیار برابر با عدد یک خواهد بود.

$$\begin{aligned} Conf(A \rightarrow B) &= \frac{SUP(A \cup B)}{SUP(A)} \\ Lift(A \rightarrow B) &= \frac{Conf(A \rightarrow B)}{SUP(B)} \\ Conv(A \rightarrow B) &= \frac{1 - SUP(B)}{1 - Conf(A \rightarrow B)} \end{aligned} \quad (۲)$$

طبق عبارت (۳)، معیار Leverage دارای مقداری بین $-۰/۲۵$ و $+۰/۲۵$ می‌باشد. ایده پنهان در این معیار آن است که اختلاف بین میزان هم‌اتفاقی سمت چپ و راست قانون با آن مقداری که موردانتظار است، به چه اندازه می‌باشد.

$$Leve(A \rightarrow B) = SUP(A \cup B) - SUP(A) \times SUP(B) \quad (۳)$$

جهت توسعه نتایج این پژوهش، پیشنهاد می‌شود از روش‌های دیگر خوشه‌بندی مانند SOM و خوشه‌بندی فازی و همچنین از دیگر روش‌های بهینه‌سازی تعداد خوشه‌ها مانند الگوریتم فرابتکاری ژنتیک و ازدحام ذرات استفاده شود.

۷) پیشنهادات مدیریتی

با استفاده از نتایج به‌دست آمده بیماران بستری‌شده در اورژانس، مورد جراحی قرار نمی‌گیرند و طول مدت اقامت کمتری دارند. بنابراین مدیریت بیمارستان باید تلاش خود را در مدیریت بیماران با طول اقامت بیشتر، معطوف سایر بخش‌ها نماید. بیماران بستری‌شده در اورژانس، به تزریق خون نیازی ندارند لذا مدیریت بانک خون جهت رزرو گروه‌های مختلف خون باید توجه بیشتری به اتاق‌های جراحی داشته باشد. محدودیت‌های تحقیق را می‌توان در تک بیمارستان بودن محل مطالعه تحقیق، عدم ثبت داده‌های مربوط به سوابق بیماران و عدم وجود سنسورهای دقیق اندازه‌گیری برخی از پارامترها مانند عوامل محیطی عنوان نمود.

منابع

- Amerioun A, Teymourzadeh E, Alijanzadeh M, Hakimzadeh SM. (2014). Research Needs Assessment and Priority Setting for Health Economics : A Mixed method Study in Iran. 16(1):23–8. <https://doi.org/10.1017/S1744133105001040>
- Angadi G, Chen R. (2007). Optimization of patient length of stay at Bridgepoint Hospital. BSc Thesis] Toronto, University of Toronto. <https://doi.org/10.1007/s001340050880>
- Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S. (2018). Analysis of length of hospital stay using electronic health records : A statistical and data mining approach. 1–16. <https://doi.org/10.1371/journal.pone.0195901.t004>
- Dol SM, Jawandhiya PM. (2023). Classification technique and its combination with clustering and association rule mining in educational data mining—A survey. Engineering applications of artificial intelligence. Jun 1;122:106071. <https://doi.org/10.1016/j.engappai.2023.106071>

- Fakhri Nazari, M. Fakhri Nazari, E. & Bakhsh, N. (2023). Presenting a method to reduce the sensitivity of incremental clustering algorithms of XML documents based on collective intelligence algorithms. *Enginiiring Management and Soft Computing*, 9 (2). 177-187. <https://doi.org/10.22091/jemsc.2024.9990.1176>
- Ghadri N. (2016). Data mining with data algorithms in health trends. *J Sci Eng Elit*. (In Persian) <https://doi.org/10.1177/1460458216678443>
- Hosseini-Shokouh SM, Sadeghian K, Ameryoun A, Zaboli R. (2019) Organizational factors affecting a patient's average length of stay in hospital: Systematic review. *J Mil Med*. 20(6):577–88. <http://doi.org/10.1097/BOT.0000000000000221>
- Khajehali N, Alizadeh S. (2017). Extract critical factors affecting the length of hospital stay of pneumoni patient by data mining (case study: an Iranian hospital). *Artif Intell Med*. <https://doi.org/10.1016/j.artmed.2017.06.010>
- Mohammadkarim B, Jamil S, Pejman H, Seyyed MH, Mostafa N. (2011). Combining multiple indicators to assess hospital performance in Iran using the Pabon Lasso Model. *Australas Med J*. 4(4):175. <https://doi.org/10.4066/AMJ.2011.620>
- Molae Fard, R. (2023). Provide a method to diagnose and optimize diabetes using data mining methods and firefly algorithm. *Enginiiring Management and Soft Computing*, 9 (1). 36-48 <https://doi.org/10.22091/JEMSC.2022.6575.1147>
- Nasiripour AA, Riyahi L, Gholamipour MA, others. (2010)The effect of the presence of full time gynecologist on length of stay among inpatient of gynecology ward in yazd social security hospital, 2008. *J Med Councl Islam Repub Iran*. 28(2):169–235. <https://doi.org/10.1016/j.phrp.2016.04.008>
- Saeidi-Mobarakeh, Z. & Amoozad Khalili, H. (2024). Nonlinear Multi-Objective Optimization Model for Production Planning Based on Fuzzy Logic and Machine Learning. *Enginiiring Management and Soft Computing*, 10 (1). 155-189. <https://doi.org/10.22091/jemsc.2024.11186.1197>
- Siddique, S.M., Tipton, K., Leas, B., Greysen, S.R., Mull, N.K., Lane-Fall, M., McShea, K. and Tsou, A.Y., (2021). Interventions to reduce hospital length of stay in high-risk populations: a systematic review. *JAMA network open*, 4(9), pp.e2125846-e2125846. [doi:10.1001/jamanetworkopen.2021.25846](https://doi.org/10.1001/jamanetworkopen.2021.25846)
- Siskou O, Kaitelidou AED. (2009). Private expenditure and the role of private health insurance in Greece : status quo and future trends. 467–74. <https://doi.org/10.1007/s10198-009-0164-3>
- Tayefi M, Tajfard M, Saffar S, Hanachi P, Reza A, Esmaeily H, et al. (2017). Computer Methods and Programs in Biomedicine hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Comput Methods Programs Biomed*. 141:105–9. <https://doi.org/10.1016/j.cmpb.2017.02.001>
- Yaghoubi M, Agha Rahimi Z, Javadi M. (2013). Factors affecting in productivity index in Noor & Ali Asghar hospital in Esfahan base on hierarchical analysis process technique. *Heal Inf Manag*. 10(4):600. <https://doi.org/10.22122/him.v10i4.657>
- Yang H, Chen YP. (2015). Expert Systems with Applications Data mining in lung cancer pathologic staging diagnosis : Correlation between clinical and pathology information. *Expert Syst Appl*. 42(15–16):6168–76. <https://doi.org/10.1016/j.eswa.2015.03.019>
- Yaya, L.H.P., Fortià, M.F., Canals, C.S. et al. (2015). Service quality assessment of public transport and the implication role of demographic characteristics. *Public Transp* 7, 409–428. <https://doi.org/10.1007/s12469-014-0099-7>