



# Identifying of Pattern Outpatient Visit of Insured Person with Data Mining Techniques: Case Study of Health Insurance in Bushehr Province

Rajabali Daroudi<sup>1</sup> PhD, Ebrahim Jaafaripooyan<sup>2</sup> PhD, Houshang Golzar<sup>1\*</sup> MA, MPH

<sup>1</sup> National Center for Health Insurance Research, Tehran, Iran

<sup>2</sup> Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

\*Correspondence to: Houshang Golzar, Email: golzarbs@gmail.com

Received: September 28, 2023

Revised: November 29, 2023

Accepted: January 14, 2024

Online Published: January 31, 2024

## Abstract

**Introduction:** With the advent of new challenges in health insurance industry, there has been a heightened effort to find innovative solutions to enhance service quality, optimize resource management, and increase the satisfaction of insured individuals. One significant approach in improving this domain involves the application of data mining techniques to identify behavioral patterns among health insurance policyholders during outpatient visits to diagnostic and treatment facilities.

**Methods:** The present study is a descriptive cross-sectional study. The claim data of health insurance in Bushehr province of Iran was used. After data preparation, analysis was performed using SPSS Clementine 12.0 software. The values of insurance start time, number of visits, and the value of the type of insurance were used to model the K-means algorithm in two modes including demographic mode and Recency-frequency-monetary (RFM). Sampling was done by census method. The statistical population includes the information of all outpatient referrals of the insured covered by health insurance of Bushehr province to 1,420,579 referrals to diagnostic and medical centers in 2018, which has been prepared by the researcher's direct referral to the database of medical records.

**Results:** The root mean square deviation values for RFM-based clustering and demographics are 21 and 21.65, respectively. And the Dunn's Index confirmed the better RFM-based clustering. The RFM-based K-Means algorithm classified the data into four clusters, with 44% of the insured in Cluster One, 4% in Cluster Two, 22% in Cluster Three, and 30% in Cluster Four. Based on this, cluster 2 insured, including women with insurance of other classes with 4% of the population, were identified as the most referred, and cluster 3, including women with rural insurance, with 22% of the population, were identified as the least referred insured.

**Conclusion:** The obtained model divided the insured into 4 clusters. This model allows the organization to predict the referral patterns of each insurer based on their age, gender, and type of insurance and provide appropriate services for different clusters. By using these models and technique in decision making process, the insurers satisfaction will be improved.

**Keywords:** Data Mining, Clustering, Health Insurance

## Highlights

1. Identifying the behavioral pattern of the insured's referrals can lead to the formulation of more effective strategies and data-based decisions to improve the performance of the organization.
2. By analyzing the clustering of patients, in addition to predicting treatment costs, it is also possible to group diagnostic treatment centers of the contracting parties with the aim of optimizing services.

## Citation:

Daroudi R, Jaafaripooyan E, Golzar H. Identifying of pattern outpatient visit of insured person with data mining techniques: case study of health insurance in Bushehr province. Iran J Health Insur. 2024;6(4):280-7.



# شناسایی الگوی رفتار بیمه‌شدگان در مراجعات سرپایی با تکنیک‌های داده‌کاوی: مطالعه موردی بیمه سلامت استان بوشهر

رجبعلی درودی<sup>۱</sup> PhD، ابراهیم جعفری پویان<sup>۲</sup> PhD، هوشنگ گلزار<sup>۱\*</sup> MA, MPH

<sup>۱</sup> مرکز ملی تحقیقات بیمه سلامت، تهران، ایران

<sup>۲</sup> گروه مدیریت، سیاست‌گذاری و اقتصاد سلامت، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

\* نویسنده مسئول: هوشنگ گلزار، پست الکترونیک: golzarbs@gmail.com

انتشار آنلاین: ۱۴۰۲/۱۱/۱۱

پذیرش: ۱۴۰۲/۱۰/۲۴

تصحیح: ۱۴۰۲/۰۹/۰۸

دریافت: ۱۴۰۲/۰۶/۰۶

## چکیده

**مقدمه:** با ظهور چالش‌های جدید صنعت بیمه سلامت، تلاش برای یافتن راهکارهای نوین برای بهبود کیفیت خدمات، مدیریت بهینه منابع و افزایش رضایت بیمه‌شدگان اهمیت بیشتری پیدا کرده است. یکی از رویکردهای مهم در بهبود این حوزه، استفاده از تکنیک‌های داده‌کاوی به منظور شناسایی الگوهای رفتاری بیمه‌شدگان در مراجعات سرپایی به موسسات تشخیصی و درمانی است.

**روش بررسی:** مطالعه حاضر از نوع توصیفی-مقطعی بوده و مجموعه داده‌های آن متعلق به اداره کل بیمه سلامت استان بوشهر است. پس از آماده‌سازی داده‌ها، تحلیل با استفاده از ویرایش ۱۲ نرم‌افزار SPSS Clementine انجام شد. برای مدل‌سازی الگوریتم K-means در دو حالت جمعیت‌شناختی و مبتنی بر RFM شامل مقادیر زمان شروع بیمه، تعداد دفعات مراجعه و ارزش نوع بیمه اجرا شد تا ارتباط بین متغیرهای مختلف به دست آید. نمونه‌گیری با روش سرشماری انجام شد. جامعه آماری شامل اطلاعات کلیه مراجعات سرپایی بیمه‌شدگان تحت پوشش بیمه سلامت استان بوشهر به تعداد ۵۷۹،۴۲۰،۱ مراجعه در سال ۱۳۹۷ به مراکز تشخیصی درمانی است که با مراجعه مستقیم پژوهشگر به بانک اطلاعاتی اسناد پزشکی تهیه شده است.

**یافته‌ها:** مقدار شاخص ریشه میانگین مربع انحراف از معیار برای خوشه‌بندی مبتنی بر RFM و جمعیت‌شناسی به ترتیب ۲۱ و ۲۱/۶۵ است. و شاخص دان (Dunn) بهتر بودن خوشه‌بندی مبتنی بر RFM را تأیید کرد. الگوریتم K-Means مبتنی بر RFM، داده‌ها را در ۴ خوشه قرار داد، به طوری که ۴۴ درصد بیمه‌شدگان در خوشه یک، ۴ درصد در خوشه دو، ۲۲ درصد در خوشه سه و ۳۰ درصد در خوشه چهار قرار گرفتند. بر این اساس بیمه‌شدگان خوشه ۲ شامل زنان دارای بیمه سایر اقسار با ۴ درصد جمعیت به عنوان پرمراجعه‌ترین و خوشه ۳ شامل زنان دارای بیمه روستایی با ۲۲ درصد جمعیت کم‌مراجعه‌ترین بیمه‌شدگان مشخص شدند. **نتیجه‌گیری:** مدل به دست آمده بیمه‌شدگان را در ۴ خوشه قرار داد. این تحلیل به سازمان این امکان را می‌دهد که الگوهای مراجعه برای هر یک از این خوشه‌ها را نسبت به متغیرهای سن، جنسیت و نوع بیمه شناسایی کند. نقطه قابل بهبود در این حیطه حفظ منافع بیمه‌شدگان و ارائه خدمات شایسته و درخور برای هر خوشه است. نتایج نهایی حاصل شده نشان‌دهنده وجود الگوهای مطلوب به عنوان یک ابزار تصمیم‌گیری مناسب برای افزایش رضایت‌مندی بیمه‌شدگان در سازمان مطرح می‌شود.

**واژگان کلیدی:** داده‌کاوی، خوشه‌بندی، بیمه سلامت

## نکات ویژه

- ۱- شناسایی الگوی رفتاری مراجعات بیمه‌شدگان می‌تواند به تدوین استراتژی‌های اثربخش‌تر و تصمیم‌گیری‌های مبتنی بر داده برای بهبود عملکرد سازمان، منجر شود.
- ۲- با تحلیل خوشه‌بندی بیماران، علاوه بر پیش‌بینی هزینه‌های درمان، می‌توان با هدف بهینه‌سازی خدمات، مراکز تشخیصی درمانی طرف قرارداد را نیز گروه‌بندی کرد.

## مقدمه

جامعه بیشتر می‌شود. یکی از اهداف سازمان بیمه سلامت، کاهش هزینه‌ها و افزایش کیفیت خدمات سلامت است. از متغیرهای مهم و تأثیرگذار بر این هدف، داده‌های سلامت است. داده‌ها در عصر امروزی، عمده‌ترین

حوزه سلامت بنا بر آمارها چهارمین حوزه بزرگ از لحاظ مالی در دنیاست و به‌طور مستمر در حال تولید میزان زیادی از داده‌هاست. به‌خصوص با پیشرفت جوامع توجه به سلامت افراد

شده است. بر طبق CRISP-DM یک پروژه داده کاوی معین، شامل چرخه حیات شش مرحله‌ای مطابق شکل ۱ است.

الگوی RFM اولین بار توسط هوگس در سال ۱۹۹۴ معرفی شد و برای بخش‌بندی مشتریان در علم بازاریابی به کار می‌رود. او برای تحلیل RFM، از رفتار گذشته مشتری که به آسانی قابل پیگیری و دسترسی است، استفاده کرد. این الگو از سه بعد مربوط به داده‌های مبادلاتی مشتریان، برای تحلیل رفتار آنها استفاده می‌کند. شاخص‌های این الگو به صورت زیر تعریف می‌شوند [۵]. شاخص تازگی مبادله (Recency) بر فاصله زمانی بین آخرین خرید صورت گرفته توسط مشتری تا پایان دوره خاص (پایان محدوده زمانی مورد بررسی) اشاره دارد. کمتر بودن این فاصله نشانگر بالا بودن ارزش این شاخص در الگو است. شاخص تعداد تکرار مبادله (Frequency) بیانگر تعداد مبادلاتی است که یک مشتری در یک دوره زمانی خاص انجام داده است. بیشتر بودن تعداد مبادلات، نشانگر بالا بودن ارزش این شاخص در الگو است. شاخص ارزش پولی مبادله (Monetary) نشان‌دهنده مقدار پولی است که یک مشتری در یک دوره زمانی خاص برای مبادلات، صرف کرده است. بیشتر بودن مقدار پول صرف شده، بیانگر بالا بودن ارزش این شاخص در الگو است.

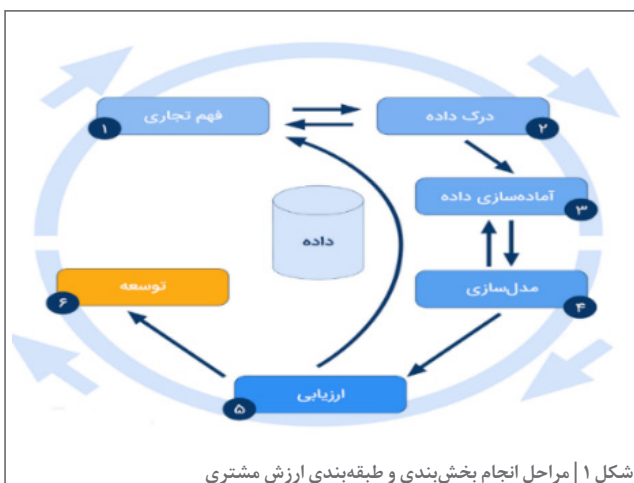
با مرور پژوهش‌های انجام شده برای شناسایی رفتار مشتریان می‌توان دریافت که تکنیک غالب در این پژوهش‌ها استفاده از الگوریتم‌های خوشه‌بندی K-Means با استفاده از الگوی RFM است. نظر به اینکه در مطالعات داخلی تا کنون داده‌های سازمان بیمه سلامت به عنوان بزرگترین سازمان بیمه‌گر پایه، با استفاده از الگوهای داده کاوی مورد بررسی قرار نگرفته، مطالعه حاضر می‌تواند سازمان را در شناخت توان داده‌های موجود در بانک‌های

دارایی سازمان‌های سلامت بوده و موفقیت سازمان‌های سلامت در گرو جمع‌آوری، ذخیره و تحلیل آنهاست. برای تبدیل این ارزش بالقوه به اطلاعات استراتژیک، بسیاری از سازمان‌ها به داده کاوی روی آورده‌اند زیرا به واسطه داده کاوی امکان کشف روابط، روندها و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین در زمینه چالش‌های آشکار و نهان سازمان میسر خواهد شد. تصمیم‌گیری در خصوص خرید خدمات سلامت و پوشش خدمات از حالت سنتی به مبتنی بر شواهد از نیازهای سازمان بیمه سلامت در استفاده از داده کاوی است. مطالعات متعددی مؤید این است که تکنیک‌های داده کاوی ابزار مؤثری را برای شناسایی الگوهای مهم سلامت از درون پرونده‌های پزشکی فراهم می‌کند [۱].

یکی از حوزه‌هایی که در این زمینه به سرعت در حال رشد است، سیستم‌های داده کاوی در مدیریت ارتباط با مشتری است و یکی از پرکاربردترین روش‌های داده کاوی در بحث پیش‌بینی رفتار مشتریان، خوشه‌بندی است. خوشه‌بندی یا بخش‌بندی مشتریان شامل تقسیم‌بندی مشتریان به گروه‌های اصلی است که شباهت بیشتری به یکدیگر دارند [۲].

به طور کلی تحلیل داده‌های توصیفی کامپیوتری، در مجموعه‌های بزرگ و پیچیده داده‌ها را داده کاوی نامیده‌اند [۳]. داده کاوی یک حوزه میان‌رشته‌ای با رشد سریع است که حوزه‌های مختلفی همچون پایگاه داده، آمار (مطالعه عددی روابط داده‌ها)، هوش مصنوعی (هوش انسان مانند نرم‌افزار و/یا ماشین) و یادگیری ماشین (الگوریتم‌هایی است که می‌توانند از داده‌ها برای پیش‌بینی استفاده کنند) و سایر زمینه‌های مرتبط را با هم تلفیق کرده تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده‌ها را استخراج کند. داده کاوی، اکتشاف و تحلیل حجم زیادی از داده‌ها برای کشف الگوها و قواعد معنادار است. فرایند داده کاوی گاهی «کشف دانش» نیز نامیده می‌شود [۴].

پروژه‌های داده کاوی، در چارچوب فرایند استاندارد اجرا می‌شود که توسط یک تیم کاری در شرکت SPSS در قالب پروژه‌ای به نام «فرایند اجرای داده کاوی استاندارد صنعتی» CRISP-DM ارائه شده است. روش‌های مختلفی برای پیاده‌سازی و اجرای پروژه‌های داده کاوی وجود دارد. یکی از روش‌های بسیار قوی، روش‌شناسی استاندارد داده کاوی CRISP-DM است. این روش اولین بار در سال ۱۹۹۶ ارائه شده است. تحقیقی که سال ۲۰۱۰ انجام شد، نشان می‌دهد که این روش‌شناسی بیش از ۵۰ درصد نسبت به سایر روش‌شناسی‌ها در سال ۲۰۰۷ به کار گرفته

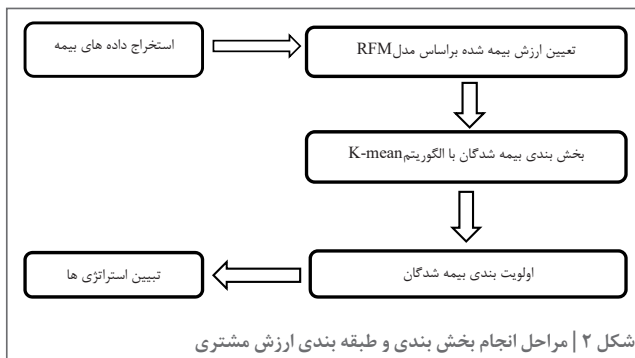


شکل ۱ | مراحل انجام بخش‌بندی و طبقه‌بندی ارزش مشتری

شبهات را به یکدیگر و بیشترین تفاوت را با اعضای خوشه‌های دیگر داشته باشند [۱۲].

پژوهش حاضر بر اساس روش‌شناسی استاندارد داده‌کاوی CRISP-DM انجام شده است [۶]. ویژگی‌های آماری رکوردهای تحت بررسی شامل نوع صندوق بیمه، جنس، تاریخ تولد، تاریخ شروع بیمه، و تعداد مراجعه‌شدگان است که در جدول ۱ نشان داده شده است. ملاحظه می‌شود که ۱۴۸۹۲۹ نفر بیمه شده در مجموع ۱۱۲۱۷۱۳ مراجعه داشته‌اند و بیشترین درصد مراجعات در صندوق بیمه‌ای کارکنان دولت با ۳۸ درصد و پس از آن روستاییان با ۳۵ درصد است. همچنین ۵۶ درصد مراجعه‌کنندگان زن و ۴۴ درصد مرد هستند.

استفاده از شاخص RFM به دلیل رایج بودن این ابزار برای اندازه‌گیری اهمیت مشتری در کسب و کارها صورت می‌گیرد. RFM



شکل ۲ | مراحل انجام بخش بندی و طبقه بندی ارزش مشتری

جدول ۱ | ویژگی‌های جمعیت‌شناختی داده مورد بررسی

نوع متغیر	انحراف معیار ± میانگین	درصد °	تعداد	مشخصه
---	۳۸	۵۶۰۸۲	کارکنان دولت	نوع صندوق بیمه
---	۷	۱۰۲۴۹	سایر اقشار	
---	۳۵	۵۲۸۰۸	روستاییان	
---	۲	۲۳۷۲	ایرانیان	
---	۱۸	۲۷۴۱۸	بیمه سلامت همگانی	بیمه سلامتی همگانی
---	۵۶	۸۳۹۳۸	زن	
---	۴۴	۶۴۹۹۱	مرد	جنس
فاصله‌ای	---	---	۱۴۸۹۲۹	تاریخ تولد
فاصله‌ای	---	---	۱۴۸۹۲۹	تاریخ شروع بیمه
عددی	± ۱۳/۱۷ ۱۰/۹۶	---	۱۱۲۱۷۱۳	تعداد مراجعه

\* متغیرهای عددی به صورت «انحراف معیار ± میانگین» و متغیرهای اسمی و ترتیبی به صورت «درصد» گزارش شده‌اند.

اطلاعاتی در راستای استفاده از این داده‌های بزرگ توانمند سازد. استفاده از این الگوها در برای خرید راهبردی خدمات سلامت برای بیمه‌شدگان توسط سازمان باعث کاهش هزینه‌ها و افزایش رضایت‌مندی می‌شود.

## روش بررسی

مطالعه حاضر از نوع توصیفی-تحلیلی و مقطعی است. در این مطالعه نمونه‌گیری با روش سرشماری انجام شد. جامعه آماری شامل اطلاعات کلیه مراجعات سرپایی بیمه‌شدگان تحت پوشش بیمه سلامت استان بوشهر به تعداد ۱,۴۲۰,۵۷۹ مراجعه به مراکز تشخیصی درمانی است که با مراجعه مستقیم پژوهشگر به بانک اطلاعاتی اسناد پزشکی تهیه شده است. ابزار جمع‌آوری داده‌ها، بانک اطلاعاتی اسناد پزشکی بوده و داده‌ها با استفاده از جستجو و اخذ گزارش از بانک اطلاعاتی مذکور استخراج و به صورت اکسل ذخیره شدند. در سال ۱۳۹۷ جمعیت بیمه شده تحت پوشش اداره کل بیمه سلامت استان بوشهر ۴۴۹,۸۲۸ نفر بود که در مجموع ۱,۴۵۳,۹۸۹ مراجعه به مراکز تشخیصی درمانی طرف قرارداد داشته‌اند. ۱,۴۲۰,۵۷۹ (۹۸ درصد) از این مراجعات به صورت سرپایی بوده که ۴۳ درصد هزینه‌های درمانی را تشکیل می‌دهد. در مطالعه حاضر برای سنجش میزان استفاده بیمه‌شدگان از خدمات سلامت بر اساس اطلاعات جمعیت‌شناختی، الگوهای مراجعه سرپایی را با استفاده از خوشه‌بندی در دو حالت جمعیت‌شناختی و مبتنی بر (RFM) Monetary (Recency-Frequency) شامل مقادیر زمان شروع بیمه، تعداد دفعات مراجعه و ارزش نوع بیمه تبیین شد. تشخیص الگو روندی است که توسط آن، مشاهدات مختلف در یک مطالعه طبقه‌بندی و توصیف می‌شوند. در حقیقت دریافت داده‌های خام و تصمیم‌گیری بر اساس دسته‌بندی داده‌هاست. مراحل انجام کار در این مطالعه مبتنی بر RFM و تکنیک‌های داده‌کاوی بر اساس شکل ۲ است [۶].

تحلیل داده‌ها با استفاده از ویرایش ۱۲ نرم‌افزار SPSS Clementine انجام شد. برای مدل‌سازی الگوریتم K-means در دو حالت جمعیت‌شناختی و مبتنی بر RFM شامل مقادیر زمان شروع بیمه، تعداد دفعات مراجعه و ارزش نوع بیمه اجرا شد تا ارتباط بین متغیرهای مختلف به دست آید. الگوریتم K-Means روش خوشه‌بندی برای گروه‌بندی تعدادی از پیشینه‌هاست. این گروه‌بندی بر اساس شباهت مقدار ورودی برای مجموعه داده‌هاست. به طوری که نمونه‌های هر خوشه (به تعداد k خوشه) بیشترین

بهرتر داده‌هاست [۱۲].

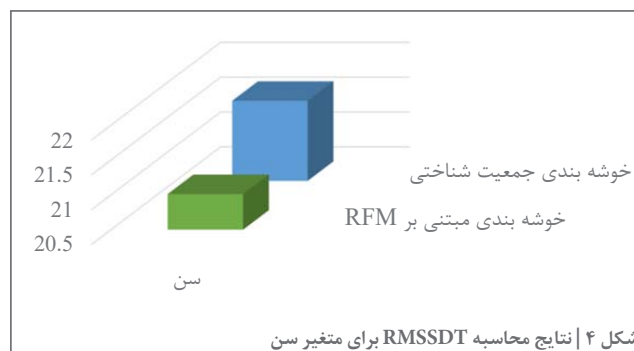
$$RMSSDT = \sqrt{\frac{\sum_{i=1 \dots n_c} \sum_{j=1 \dots d} n_{ij} (x - \bar{x}_j)^2}{\sum_{i=1 \dots n_c} (n_{ij} - 1)}} \quad (4)$$

### یافته‌ها

پس از اجرای خوشه‌بندی K-Means مبتنی بر RFM و جمعیت‌شناختی روی داده‌ها، برای به‌دست آوردن تعداد بهینه خوشه از شاخص دان استفاده شد. شاخص دان نشان داد که عدد خوشه بهینه برای خوشه‌بندی جمعیت‌شناختی ۹ و برای خوشه‌بندی مبتنی بر RFM ۴ است (شکل ۳).

در مدل‌سازی حاضر به دلیل استخراج سن از روی سال تولد، شاخص ریشه میانگین مربع انحراف از معیار برای این متغیر عددی در هر دو مدل محاسبه شد. نتیجه نشان داد که خوشه‌بندی مبتنی بر RFM بهترین مدل را تولید کرده، زیرا مقدار شاخص ریشه میانگین مربع انحراف از معیار برای متغیر سن در این روش کمتر است (شکل ۴).

بنابراین الگوریتم K-Means مبتنی بر RFM، داده‌ها را در ۴ خوشه قرار داد به طوری که ۴۴ درصد بیمه‌شدگان در خوشه یک، ۴ درصد در خوشه دو، ۲۲ درصد در خوشه سه و ۳۰ درصد در خوشه چهار قرار گرفتند. جدول ۳ نتایج حاصل از خوشه‌بندی مبتنی بر RFM را نشان می‌دهد. در خوشه یک تمام مردان قرار



شامل مقادیری است که میزان وفاداری مشتری را به محصولات یک کسب و کار نشان می‌دهند. مقادیر RFM را می‌توان به طور مناسب و مخصوص کسب و کار تعریف کرد. بدین ترتیب مقایسه RFM در بیمه سلامت با متغیرهای تاریخ شروع بیمه (R)، تعداد مراجعه به مراکز تشخیصی درمانی (F) و ارزش نوع بیمه (M) برای هر بیمه شده (رکورد) جایگزین شدند. ارزش نوع بیمه بر اساس مدت اعتبار و دسترسی به مراکز طرف قرارداد صورت گرفت. برای بیمه کارکنان دولت که اعتباردهی آن ۱۰ ساله بوده و دسترسی به همه مراکز (دانشگاهی، غیردانشگاهی و خصوصی) برای آن وجود دارد، بالاترین ارزش و برای بیمه سلامت همگانی که اعتباردهی آن یک ساله و دسترسی آن فقط به مراکز دانشگاهی است، پایین‌ترین ارزش مطابق جدول ۲ در نظر گرفته شد.

برای ارزیابی مدل و انتخاب عدد خوشه بهینه از شاخص دان (Dunn) (رابطه ۱) استفاده شد. هدف شاخص دان حداکثرسازی فاصله برون خوشه‌ای در ضمن حداقل کردن فاصله‌ی درون خوشه‌ای است تا نمونه‌های مشابه در یک خوشه قرار گیرند.

$$D = \min_{i=1 \dots n_c} \left\{ \min_{j=i+1 \dots n_c} \left( \frac{d(c_i, c_j)}{\max_{k=1 \dots n_c} (\text{diam}(c_k))} \right) \right\} \quad (1)$$

که  $d(c_i, c_j)$  و  $\text{diam}(c_i)$  طبق روابط (۲) و (۳) محاسبه می‌شوند.

$$d(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(x, y)\} \quad (2)$$

$$\text{diam}(c_i) = \max_{x, y \in c_i} \{d(x, y)\} \quad (3)$$

برای ارزیابی مدل‌ها و انتخاب مدل بهتر از شاخص ریشه میانگین مربع انحراف از معیار استفاده شده است (رابطه ۴). این معیار دقت مدل را بر اساس تفاضل بین ارزش واقعی و ارزش برآورد شده ارزیابی می‌کند و هرچه کمتر باشد نشان‌دهنده خوشه بندی

جدول ۲ | ارزش‌گذاری نوع بیمه

نوع بیمه	اعتبار دهی	دانشگاهی	غیردانشگاهی	دسترسی به مراکز طرف قرارداد خصوصی	ارزش متغیر M
کارکنان دولت	۱۰ سال	دارد	دارد	دارد	۵
سایر اقشار	۱ سال	دارد	دارد	دارد	۴
روستاییان	۵ سال	دارد	ندارد	ندارد	۳
ایرانیان	۱ سال	دارد	دارد	دارد	۲
بیمه سلامت همگانی	۱ سال	دارد	ندارد	ندارد	۱

جدول ۳ | نتایج حاصل از خوشه‌بندی مبتنی بر RFM

خوشه ۴۴۶۹۹۴ عضو	خوشه ۳۳۱۱۰۳ عضو	خوشه ۶۱۲۹۲ عضو	خوشه ۶۴۹۹۱۱ عضو	مشخصه	
			۴۱/۳	کارکنان دولت	نوع صندوق
		۱۰۰	۶/۳۴	سایر اقشار	
	۱۰۰		۳۰/۳۱	روستاییان	
۲/۵۳			۱/۹۱	ایرانیان	
۳۲/۰۶			۲۰/۱۴	بیمه سلامت همگانی	
۱۰۰	۱۰۰	۱۰۰		زن	جنس
			۱۰۰	مرد	
۲۶/۸۰۳ ± ۱۹/۶۱۹	۳۷/۰۴۳ ± ۱۹/۵۴۹	۴۵/۸۴۸ ± ۲۱/۶۷۹	۳۵/۴۲۷ ± ۲۲/۵۳۷ °	سن	سن
۱۰۳/۷۹۶ ± ۹۳/۹۲۹	۱۵۱/۳۹۵ ± ۱۰۴/۴۲۱	۷۴/۳۵۶ ± ۸۰/۴۱	۱۲۷/۳۱۷ ± ۱۰۲/۱۵	شروع بیمه	شروع بیمه
۱۴/۱۶۹ ± ۱۴/۰۵۴	۵/۸۷۷ ± ۶/۹۶۴	۲۴/۱۳۵ ± ۲۳/۰۶۲	۱۰/۱۱۲ ± ۱۲/۲۳۶	تعداد مراجعه	تعداد مراجعه
۵۸/۰۴۷ ± ۷۱/۱۵۹	۱۷/۶۳۱ ± ۲۰/۸۹۱	۹۶/۵۳۹ ± ۹۲/۲۴۷	۳۹/۷ ± ۵۷/۱۴۵	ارزش نوع بیمه	ارزش نوع بیمه

\* متغیرهای کمی به صورت «انحراف معیار ± میانگین» و متغیرهای اسمی به صورت «درصد» گزارش شده‌اند.

در یک خوشه قرار گرفته‌اند. این موضوع در خصوص زنان روستایی، با توجه به نظام ارجاع، میزان مراجعه آنها با زنان سایر صندوق‌های بیمه‌ای متفاوت بوده و در خوشه‌ای مجزا قرار دارند.

جدول ۴ رتبه هر خوشه را با توجه به مقادیر شروع بیمه، تعداد مراجعه و ارزش بیمه پس از بررسی رفتار بیمه‌شدگان نشان می‌دهد. رتبه‌بندی از پرمراجعه به کم‌مراجعه صورت گرفته است. هنگام تخصیص رتبه به هر خوشه باید در نظر داشت هرچه مقدار متغیر شروع بیمه کمتر و دفعات مراجعه و ارزش بیمه بیشتر باشد رتبه بهتر می‌شود. براساس رتبه‌بندی صورت گرفته خوشه ۲ شامل زنان دارای بیمه سایر اقشار رتبه ۱ میزان مراجعه را دارند و عموماً والدین شهدا و مددجویان بهزیستی تحت پوشش بیمه‌ای این صندوق قرار دارند. خوشه ۴ شامل زنان دارای بیمه کارکنان دولت، ایرانیان و بیمه سلامت همگانی دارای رتبه ۲ میزان مراجعه هستند. همچنین خوشه ۱ شامل تمام مردان در کلیه صندوق‌های بیمه‌ای در رتبه ۳ میزان مراجعه بوده و در نهایت خوشه ۳ شامل زنان روستایی در رتبه ۴ میزان مراجعه قرار دارند.

جدول ۴ | رتبه‌بندی بیمه‌شدگان براساس مقادیر RFM

نام خوشه	شروع بیمه	تعداد مراجعه	ارزش بیمه	رتبه
خوشه ۱	۱۲۷/۳	۱۰/۱	۳۹/۷	۳
خوشه ۲	۷۴/۳	۲۴/۱	۹۶/۵	۱
خوشه ۳	۱۵۱/۳	۵/۸	۱۷/۶	۴
خوشه ۴	۱۰۳/۷	۱۴/۱	۵۸/۰	۲

دارند و ترکیبی از همه صندوق‌های بیمه‌ای است. میانگین سنی این خوشه ۳۵ سال و میانگین دفعات مراجعه آنان ۱۰ است. همچنین در این خوشه بیشترین بار مراجعات مربوط به رده سنی زیر ۶ سال است. در خوشه دو که ۱۰۰ درصد آن زنان صندوق بیمه سایر اقشار قرار دارند و میانگین سنی این خوشه تقریباً ۴۶ سال و بیشتر از سایر خوشه‌هاست. میانگین دفعات مراجعه در این خوشه ۲۴ است و خوشه‌ای پرمراجعه محسوب می‌شود. در این خوشه بیشترین بار مراجعات مربوط به رده سنی بالاتر از ۶۶ سال است. در خوشه سه ۱۰۰ درصد آن زنان صندوق بیمه روستایی قرار دارند و میانگین سنی این خوشه تقریباً ۳۷ سال و میانگین دفعات مراجعه آنان ۵/۸ است. در خوشه ۳ بیشترین بار مراجعات به رده سنی ۳۶-۳۱ تعلق دارد. در خوشه چهار زنان با نوع بیمه کارکنان دولت، ایرانیان و بیمه سلامت همگانی قرار دارند و میانگین سنی این خوشه تقریباً ۳۷ سال و میانگین دفعات مراجعه در این خوشه ۱۴ است. بیشترین بار مراجعات مربوط به رده سنی ۴۸-۴۳ است.

مشاهده می‌شود که زنان و مردان در خوشه‌های جداگانه قرار گرفته‌اند و الگوی مراجعه در مردان و زنان متفاوت است. همچنین زنان در سه خوشه جداگانه قرار دارند که زنان صندوق سایر اقشار در خوشه ۲ و زنان روستایی در خوشه ۳ و زنان صندوق ایرانیان و بیمه سلامت همگانی در خوشه ۴ قرار گرفتند. بیمه‌شدگان صندوق روستایی از نظام ارجاع استفاده می‌کنند اما میزان مراجعه مردان روستایی همانند مردان در صندوق کارکنان دولت بوده و

## بحث

تا گروه‌های مختلف شناسایی شوند [۱۲]. مطالعه‌ای که توسط Christodoulakis و Aggelis (۲۰۰۵) با عنوان «خوشه‌بندی مشتریان با استفاده از تحلیل RFM» روی مشتریان بانکداری الکترونیک انجام شد، نشان داد که هر چه نمره RFM یک مشتری بالاتر باشد آن مشتری برای بانک سودآورتر است [۱۳].

در مطالعه‌ای که توسط وو و همکاران (۲۰۰۹) با عنوان استفاده از مدل RFM و روش K-means برای تحلیل مشتریان یک شرکت تولیدکننده در تایوان انجام شد، تجزیه و تحلیل‌ها نشان داد که خوشه‌های ۵ و ۶ دو گروه مهم هستند که شرکت مجبور است منابع خود را به آنها اختصاص دهد. اما خوشه‌های ۱ و ۲ به دلیل هزینه‌های کم، شامل جیره‌بندی توسط شرکت شدند [۱۴]. مطالعه حمدی و زمیری (۲۰۱۶) با عنوان «شناسایی و بخش‌بندی مشتریان شرکت بیمه پاسارگاد» با استفاده از مدل RFM مشتریان در ۴ خوشه قرار گرفتند که خوشه اول شامل مشتریان طلایی، خوشه دوم شامل مشتریان باارزش و وفادار، خوشه سوم مشتریان ثابت و خوشه چهارم شامل مشتریان بی‌ارزش با احتمال تمديد نکردن بیمه‌نامه است [۱۵]. بنابراین خوشه‌بندی یا کلاس‌بندی یکی از تکنیک‌های مهم در حوزه داده‌کاوی است که به مدیران می‌تواند در درک بهتر از داده‌ها و بهبود فرایندها و تصمیم‌گیری‌ها کمک کند.

باتوجه به نتایج حاصل از خوشه‌بندی مشاهده می‌شود که میانگین دفعات مراجعه بین خوشه‌ها با هم اختلاف دارند. از طرفی چون اعضای خوشه‌ها از همه صندوق‌ها نیستند، برای مثال در خوشه دو، فقط صندوق سایر اقشار و در خوشه سه فقط صندوق روستاییان هستند، بنابراین الگوی مراجعه بیمه‌شدگان در صندوق‌های بیمه‌ای یکسان نیست. همچنین مشاهده می‌شود حداقل یک خوشه، به عنوان مثال خوشه دو که ۱۰۰ درصد آن زنان صندوق بیمه سایر اقشار قرار دارند و میانگین سنی آن تقریباً ۴۶ سال است، با بقیه خوشه‌ها اختلاف دارد. در هر خوشه نیز یک رده سنی وجود دارد که میزان مراجعه آن بیشتر سایر رده‌های سنی است. به عبارتی دیگر میزان مراجعه بیمه‌شدگان از نظر بازه سنی با هم اختلاف دارند. طبق نتایج حاصل از خوشه‌بندی، خوشه یک شامل مردان و خوشه‌های دو، سه و چهار شامل زنان هستند و با توجه به یکسان نبودن میزان مراجعه در خوشه‌ها، نتیجه می‌گیریم که میزان مراجعه بیمه‌شدگان از نظر جنسیت (زن و مرد) با هم اختلاف دارند. برای مثال خوشه ۲ که شامل زنان است، پرمراجعه‌ترین خوشه است.

در ارتباط با الگوهای مراجعه و ارزش مشتری با استفاده از مدل RFM مطالعات زیادی انجام شده است. در مطالعه‌ای که توسط رجبی و همکاران تحت عنوان شناسایی الگوهای رفتار مردم در اهدا خون با استفاده از الگوریتم K-Means مبتنی بر تازگی، بسامد و ارزش خون بر روی داده‌های سازمان انتقال خون تهران انجام شد، الگوریتم داده‌ها را در چهار خوشه قرار داد که خوشه چهارم شامل مردان مجرد حائز بیشترین رتبه وفاداری اهدای خون و خوشه سه شامل زنان متأهل حائز کمترین رتبه وفاداری شد [۷]. همچنین در مطالعه مظفری و همکاران تحت عنوان ارائه روشی برای بخش‌بندی مشتریان با استفاده از مدل RFM روی مشتریان بانک خصوصی در شهر شیراز انجام شد، مشتریان به سه خوشه اصلی تحت عنوان مشتریان خوب، معمولی و بد تفکیک شدند [۸].

در مطالعه‌ای که توسط زین‌العابدینی و خان‌بایی با عنوان مدل به‌کارگیری تکنیک‌های داده‌کاوی در شناسایی، بخش‌بندی و تحلیل رفتار مشتریان خدمات بانکداری الکترونیکی انجام شد، استفاده از مدل RFM در تحلیل رفتار مشتریان نشان داد که بخش‌بندی مشتریان به کمک خوشه‌بندی، محاسبه ارزش هر مشتری در خوشه‌ها و رتبه‌بندی آنها برای پیدا کردن ارزشمندترین خوشه‌ها و امتیازدهی و تعیین ارزش مشتری به عنوان ویژگی هدف در ساخت مدل‌های طبقه‌بندی میزان ارزش مشتریان کارآمد بوده است [۹]. در مطالعه‌ای که توسط فاطمه باقری و مجید زیارت‌بان تحت عنوان «افزایش سهم بازار و مدیریت ارتباط با مشتری در شرکت‌های بیمه با استفاده از خوشه‌بندی K-mean» انجام شد مدلی بر اساس خوشه‌بندی مشتریان ارائه شده و سیاست‌های جدیدی را برای بیمه‌خودرو در یک شرکت بیمه ارائه کردند. در این مدل از روش خوشه‌بندی K-Means بر مبنای ویژگی‌های RFM مشتری استفاده شده است [۱۰]. مطالعه‌ای که توسط قصابی و همکاران با عنوان «استخراج دانش از سیستم اطلاعات سازمان تأمین اجتماعی» با رویکردهای داده‌کاوی انجام شد، برای خوشه‌بندی متغیرها از الگوریتم K میانگین و همچنین به منظور بررسی مدل و تعداد خوشه‌های بهینه از شاخص دیویس بولدین استفاده شده است [۱۱]. مطالعه‌ای که توسط فاطمه باقری با عنوان «تعیین الگوی توزیع واقعه مرگ» با استفاده از تکنیک‌های داده‌کاوی در استان گلستان انجام شد ابتدا جامعه مورد بررسی به دو گروه داده‌های مربوط متوفیان زیر یکسال و بالای یکسال تقسیم شدند سپس این دو گروه داده با استفاده از روش K-means خوشه‌بندی شدند

## References

- Balib RK. Clinical Knowledge Management: Opportunities and Challenges. Hershey: Idea Group Inc (IGI); 2005. doi: 10.4018/978-1-59140-300-5.
- Ngai EW, Xiu L, Chau DC. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst Appl.* 2009;36(2):2592-602. doi: 10.1016/j.eswa.2008.02.021.
- Friedman JH. Data Mining and Statistics: What's the connection?. *Computing Science and Statistics.* 1998;29(1):3-9.
- Han J, Kamber M, Pei J. Data mining concepts and techniques third edition. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University. 2012.
- Cheng CH, Chen YS. Classifying the segmentation of customer value via RFM model and RS theory. *Expert Syst Appl.* 2009;36(3):4176-84. doi: 10.1016/j.eswa.2008.04.003.
- Seyedhosseini SM, Gholamian MR, Maleki A. A methodology based on RFM using data mining approach to assess the customer loyalty. *International Journal of Industrial Engineering and Production Management.* 2011;22(2):171-9. [Persian]
- Rajabi J, Alizade S, Ashoori M. Identifying behavioral patterns in blood donation using K-means algorithm based on recency, frequency and blood value. *Journal of Health Administration.* 2018;21(71):Pe66-Pe77.
- Gholmian M, Mozafari A. Presenting a method for customer segmentation using the RFM model under conditions of uncertainty. *Iranian Journal of Information and Communication Technology.* 2016;8(29):41-54. [Persian]
- Zeinolabedini SF, Khanbabaei M. Model of Applying Data Mining Techniques in identification, segmentation and Analysis of Customers Behaviour of Electronic Banking Services. *New Marketing Research Journal.* 2013;3(2):177-90.
- Bagheri F, Ziaratban M. Increasing market share and customer relationship management in insurance companies using k-means clustering. 11th Iranian Conference on Intelligent Systems. 2013. [Persian]. Available from: <https://civilica.com/doc/214649/>
- Ghasabi A, Hatalo A, Hashemzadeh F. Extracting knowledge from the information system of the Social Security Organization with data mining approaches: a case study of the clients of the Social Security Organization's medical centers. *The Second National Conference of Applied Research in Computer Science and Technology.* 2015. [Persian]. Available from: <https://civilica.com/doc/455065/>
- Bagheri F. Determination of the Distribution Pattern of Mortality Using Data Mining Technique in Golestan Province since 2007 to 2009. *Jorjani Biomedicine Journal.* 2015;3(2):65-79. [Persian]
- Aggelis V, Christodoulakis D. Customer clustering using rfm analysis. In *Proceedings of the 9th WSEAS International Conference on Computers.* 2005 Jul 14. p. 2.
- Wu HH, Chang EC, Lo CF. Applying RFM model and K-means method in customer value analysis of an outfitter. In *Global Perspective for Competitive Enterprise, Economy and Ecology: Proceedings of the 16th ISPE International Conference on Concurrent Engineering.* 2009. p. 665-672. Springer London. doi: 10.1007/978-1-84882-762-2\_63.
- Hamdi K, Zamiri A. Identifying and segmenting customers of Pasargad insurance company through RFM model (RFM). *Int Bu Manag.* 2016;10(18):4209-14. [Persian]

برای استفاده از یافته‌های این پژوهش در سازمان بیمه سلامت می‌توان به خوشه‌بندی برای گروه‌بندی بیماران براساس ویژگی‌های مشترک اشاره کرد. این اطلاعات می‌توانند در تصمیم‌گیری‌های مرتبط با پویای درمانی، نظارت بر بیماری‌های خاص و تخصیص منابع بهینه مؤثر باشند. همچنین مراکز تشخیصی درمانی طرف قرارداد می‌توانند براساس نیازهای مختلف جمعیت گروه‌بندی شوند. این کمک می‌کند تا خدمات بهینه‌تر و به میزان نیاز ارائه شود و تراکم ناهمگن خدمات در نقاط مختلف را کاهش دهد. با تحلیل خوشه‌بندی بیماران بر اساس ویژگی‌های مختلف، می‌توان به پیش‌بینی هزینه‌های درمان برای هر گروه از بیماران پرداخت. این اطلاعات می‌توانند به مدیران کمک کنند تا منابع مالی را بهینه‌تر تخصیص دهند و استراتژی‌های مالی مؤثرتری را اتخاذ کنند. در نهایت استفاده از خوشه‌بندی به مدیران کمک می‌کند تا به شناخت بهتری از داده‌ها دست پیدا کنند و تصمیم‌گیری‌های مبتنی بر داده را در برای بهبود عملکرد سازمان انجام دهند.

## نتیجه‌گیری

مدل به‌دست آمده در این پژوهش بیمه‌شدگان را در ۴ خوشه قرار داد. اکنون برای سازمان روشن است که بیمه‌شده‌ای با سن، جنسیت و نوع بیمه مشخص دارای چه نوع الگوی مراجعه‌ای خواهد بود و در چه خوشه‌ای جای خواهد گرفت. به این ترتیب با شناسایی الگوی رفتاری مراجعات بیمه‌شدگان می‌توان در برای استراتژی‌های ارائه خدمات سلامت به آنها، تصمیم‌گیری بهتری اتخاذ کرد. با توجه به نتایج پژوهش حاضر، وجود رویکردی سازمان‌دهی شده برای شناخت الگوی مراجعه بیمه‌شدگان به منظور کمک به مدیران استراتژیک سازمان بیمه در برای اتخاذ تصمیمات راهبردی برای هر خوشه، ضروری است. نقطه قابل بهبود در این حیطه توصیف دقیق الگوی مراجعه هر خوشه به برای حفظ منافع بیمه‌شدگان و ارائه خدمات شایسته و درخور برای هر خوشه است.

**تشکر و قدردانی:** مقاله حاضر مستخرج از پایان نامه در مقطع MPH از دانشگاه علوم پزشکی تهران است. از تمام کسانی که در این مطالعه همکاری کردند کمال تشکر و قدردانی به عمل می‌آید.

**تأیید به اخلاقی:** پژوهش حاضر توسط کمیته اخلاق دانشگاه علوم پزشکی تهران با کد اخلاقی IR.TUMS.SPH.REC.1399.153 مورد تأیید قرار گرفته است.

**تعارض منافع:** نویسندگان این مقاله هیچ‌گونه تضاد منافی را اعلام نکردند.

**سهم نویسندگان:** همه نویسندگان سهم یکسانی در تهیه و تدوین این مطالعه داشتند.

**منابع مالی:** این پژوهش مورد حمایت مالی مرکز تحقیقات بیمه سلامت قرار گرفته است.