

Comparison Classification of Data Mining Algorithms in Medical Sciences

دسته بندی تطبیقی الگوریتم های داده کاوی در علوم پزشکی

Maryam Hassanzadeh, Seyed Ali Razavi Ebrahimi

Abstract — *Today, we can use of new technology tools for restoring and saving data in huge size and so we need a new science for searching in these huge data source and finding useful and necessary results in them. Data mining is a science that is searched automatically in huge data for finding models and association rules in them where other statistical analysis can't do that. The medical science is one of sciences that need to use of these tools for analyzing their huge data and creating predictive model with new computation ways. The purpose of this research is review of rules and application area of predictive data mining in medical sciences and presenting a frame work for creating, evaluation and exploitation of data mining models in this. In this article we reviewed researches that are published recently about predictive data mining in medicine and we try to highlight important problems and summarizing data mining ways and algorithms that use in this area as a learning set. This research is shown that data mining prediction provide necessary tools for researcher and medical doctors to improve in prevention of disease, diagnosis ways and their treatment programs. And also it seems that the way of data mining in medical is mostly clustering and classification¹.*

Keywords — data mining, medical data mining algorithms, disease predictive.

داده کاوی، الگوریتم های داده کاوی در پزشکی، پیشگویی بیماری.

۱. چکیده

امروزه با پیشرفت علم و تکنولوژی و ابزارهای فناوری، توانایی بازبینی و ذخیره داده های مهم با حجم وسیع فراهم گشته است و نیاز به علمی جهت جستجو در این داده ها و دریافت نتایج مفید لازم و ضروری گشته است. داده کاوی، جستجوی خودکار منابع داده ای بزرگ، جهت یافتن الگوها و وابستگی هایی است که تحلیل های ساده آماری قادر به انجام آن نیستند. یکی از زمینه هایی که نیازمند استفاده از این ابزار ها جهت تحلیل داده های وسیع و مدل سازی پیشگویانه با روش های محاسباتی جدید است، علم پزشکی می باشد. هدف از این تحقیق بررسی نقش و دامنه ی کاربرد پیش گوینه داده کاوی در علوم پزشکی و پیشنهاد یک چهار چوب جهت ساخت، ارزیابی و بهره برداری از مدل های داده کاوی در این زمینه است. در این مقاله به بررسی کارهای مرتبط در زمینه پیشگویی داده کاوی در پزشکی که اخیراً منتشر شده اند پرداخته می شود و سعی در برجسته نمودن مسائل مهم و خلاصه سازی روش ها و الگوریتم های به کار رفته در قالب یک مجموعه آموزشی شده است. این تحقیق نشان می دهد که پیشگویی های داده کاوی ابزارهای ضروری را برای محققان و پزشکان جهت بهبود در پیشگیری از بیماریها، روشهای تشخیص و برنامه های درمانی فراهم می نماید. طبق بررسی های به عمل آمده در این تحقیق، در اکثر موارد کاوش دانش در داده های پزشکی از روش خوشه بندی و دسته بندی استفاده شده است. کلمات کلیدی:

۲. مقدمه

داده کاوی عبارت است از اکتباس یا استخراج دانش از مجموعه ای بسیار حجیم از داده ها، به بیان دیگر، داده کاوی فرایندی است که با استفاده از تکنیکهای هوشمند، دانش را از مجموعه ای از داده ها استخراج می کند که تحلیل های ساده آماری قادر به انجام آن نیستند. داده کاوی از الگوریتم های بسیار پیچیده ریاضی جهت تقسیم بندی داده ها و پیشگویی رویدادها استفاده می کند. [1]

بیشتر افرادی که که علم داده کاوی را مورد استفاده قرار می دهند افراد متخصصی در یک زمینه خاص علمی (پزشک، رادیولوژیست، مدیر فروش، تاجر و...) بوده و نه تنها به داده های خاص خود دسترسی دارند بلکه به جمع آوری آن نیز اقدام می ورزند. هدف این افراد این است که نه تنها ترجیح می دهند داده های خود را بهتر بشناسند بلکه مایلند دانش جدیدی را در رابطه با زمینه فعالیت خود کشف نمایند. مقصود صاحبان داده، حل مسائل و مشکلات با استفاده از راههای جدید و در صورت امکان بهتر می باشد. [2]

در علم پزشکی کشف و تشخیص به موقع بیماری ها می تواند از ابتلا به بسیاری از بیماریهای مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد. با پیشرفت های بیولوژیکی امروزه و توسعه تکنولوژی و استفاده از فناوری های روز و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع آوری اطلاعات دقیقتری در مورد بیماران گردند که تحلیل آنان به دلیل حجم بالای اطلاعات و متعدد بودن

¹ M. Hassanzadeh is with Department of Computer Engineering, Payame Noor University, Rasht, Iran (email: hassanzadeh@gums.ac.ir)
S.A. Razavi Ebrahimi is with Department of Computer Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran (email: ali_razavi@pnu.ac.ir)

جهت انجام این تکنیک ها، الگوریتم های بسیاری وجود دارند که از آن میان ۱۰ الگوریتم برتر پر کاربرد [3] به شرح زیر می باشند:

- 1) K-Means
- 2) C 4.5
- 3) AdaBoost
- 4) CART
- 5) EM
- 6) SVM
- 7) Page Rank
- 8) KNN
- 9) Apriori
- 10) Naïve bays

۴. تعیین عوامل موثر در ابتلا به انواع سرطان و تعیین روش

درمانی بهینه آن

نسرین اسدی و همکاران در مقاله ای تحت عنوان "با به کارگیری داده کاوی جهت تعیین عوامل موثر در ابتلا به انواع سرطان و تعیین روش درمانی بهینه در پایگاه داده سرطان بیمارستان نمازی" [4] سعی بر آن داشتند تا با بررسی انواع سرطان از طریق تکنیک ها و روش های داده کاوی به کشف وابستگی هایی جهت تعیین عوامل موثر بر سرطان پرداخته و سپس با بررسی انواع درمان از طریق الگوریتم های داده کاوی، بهترین نوع درمان برای هر یک از انواع سرطان را پیشنهاد دهند. به این منظور از الگوریتم k-means برای خوشه بندی عوامل موثر بر بیماری نظیر جنسیت، سن، سابقه ی بیماری در اقوام، مصرف سیگار یا الکل استفاده گردید و جهت تعیین بهترین درمان از میان ۳ روش درمانی شایع: شیمی درمانی، رادیوگرافی و جراحی از الگوریتم apriori استفاده نمودند. تحقیق بر روی ۱۰۴۱۱ رکورد بیمار از پایگاه داده مرکز ثبت سرطان بیمارستان نمازی شیراز سالهای ۱۳۸۲ الی ۱۳۸۶ انجام گرفت که نتایج زیر را در بر داشته است:

تشخیص و پیشگویی: عواملی نظیر سابقه ی بیماری در اقوام، مصرف سیگار یا الکل بی تاثیر نبوده ولی دلیل محکمی برای ابتلا به سرطان نیست. بهترین درمان پیشنهاد شده از میان سه نوع روش درمان شایع: در سرطان خون: انجام شیمی درمانی کافی بوده و نیازی به جراحی و رادیوتراپی نمی باشد. در سرطان سینه: انجام جراحی ضروری است. در سرطان گوارش: انجام جراحی ضروری می باشد.

۵. تحلیل داده های میکروسکوپی:

تکنولوژی ساخت DNA وسیله ای است که به محققان امکان مشاهده رفتار ژن ها در سلولهای مختلف را فراهم می آورد. [7] نه تنها بررسی رفتار ژن مهم است بلکه تاثیر آنها با ژن های دیگر نیز مهم است. بیماریهای ژنتیکی یا بسیاری از سرطان ها باعث برخی اختلالات در ژن ها می شوند. تحلیل داده های میکروسکوپی آسان نیست چون اولاً ابعاد داده های میکروسکوپی بسیار وسیع بوده و انجام آزمایشات بر روی بیماران زنده بسیار دردناک می باشد. به این علت روشهای داده کاوی با توانایی های وسیع خود در کاوش در داده های حجیم بسیار موثر بوده و در این بخش جهت مقایسه رفتار ژن ها با رشته DNA از روش خوشه بندی برای حل مشکل پردازش داده های میکروسکوپی استفاده گردیده و متد الگوریتم آماری برای اجرای روش خوشه بندی Statistical Algorithmic Method for Bicluster Analysis

موارد، مشکل می باشد و نیاز به فناوری جدیدتری می باشد که تکنولوژی های داده کاوی به کمک الگوریتم های قدرتمند خود به این مهم دست یافته است. هدف از روشهای پیشگویی داده کاوی در پزشکی بالینی ساخت یک مدل پیشگویانه است که به پزشکان کمک می کند تا روش های پیشگیری، تشخیص و برنامه های درمانی خود را بهبود بخشند. [10] در این مقاله سعی شده تا حوزه های مختلف علوم پزشکی که توسط الگوریتم های داده کاوی مورد ارزیابی قرار گرفته و در به دست آوردن نتایج مطلوب موفق بوده اند، پرداخته گردد و همچنین الگوریتم های به کار رفته در هر یک از این علوم بازگو شود تا محققین در آینده بتوانند از این توانایی ها حداکثر استفاده را ببرند.

به این منظور در بخش سوم به معرفی روش های مهم داده کاوی و الگوریتم های به کار رفته در آن پرداخته و سپس در بخش چهارم تا نهم به مرور کارهای مرتبط در زمینه پیشگویی داده کاوی در پزشکی که اخیراً منتشر شده اند می پردازیم. در بخش چهارم کاربرد روشهای داده کاوی در تعیین عوامل موثر در ابتلا به انواع سرطان و تعیین روش های درمانی بهینه آن مطرح می گردد. در بخش پنجم با توجه به گستردگی داده های آزمایشگاهی، رفتار ژن ها مورد بررسی قرار گرفته و توانایی های علم داده کاوی در این زمینه بیان می گردد. در بخش ششم تحقیقات به عمل آمده بر روی داده های بیماران اورژانس جهت تشخیص به موقع بیماری، عنوان خواهد شد. در بخش هفتم مدت زمان بستری بیماران گوارشی مورد ارزیابی قرار خواهد گرفت. در بخش هشتم حملات قلبی پیشگویی می شود. در بخش نهم نتایج داده کاوی در انبار داده بالینی طب سنتی چینی بیان خواهد شد و در نهایت طی مقایسه روشهای انجام شده در یک جدول، بهترین روشها و الگوریتم به کار رفته در حوزه های مختلف علوم پزشکی ارائه می گردد.

۳. روش های کشف دانش در داده کاوی

داده کاوی از ساخت مدل های تحلیلی، دسته بندی و پیش بینی اطلاعات و ارائه نتایج با استفاده از ابزارهای مرتبط استفاده می کند. برای اینکه الگوریتم داده کاوی بتواند عمل استخراج دانش را بخوبی انجام دهد، نیاز به یک سری پیش پردازش ها بر روی داده ها و یک سری پس پردازش ها بر روی الگوهای استخراج شده دارد. از تکنیک هایی که برای داده کاوی استفاده می شود می توان به روشهای زیر اشاره نمود:

۱. دسته بندی^۲ (تکنیک پیشگویانه): در این روش یک نمونه به یکی از چند دسته از پیش تعریف شده دسته بندی می شود.
۲. رگرسیون^۳ (تکنیک پیشگویانه): پیش بینی یک مقدار متغیر مبنی بر متغیرهای دیگر.
۳. خوشه بندی^۴ (تکنیک توصیفی): یک دسته داده را به یکی از چند خوشه نگاشت می کند. خوشه ها گروه بندیهای دسته های داده ای هستند که بر اساس شباهت برخی از معیارها بوجود می آیند.
۴. کشف قواعد وابستگی^۵ (تکنیک توصیفی): روابط وابستگی بین خصیصه های مختلف را بیان می کند.
۵. تحلیل دنباله: الگوهای دنباله ای همچون سربهای زمانی را مدل می کند.

² Classification

³ Regression

⁴ Clustering

⁵ Association Rule Discovery

(SAMBA) بوده است. (الگوریتم samba از تکنیک تئوری گراف و مدل کردن احتمالات استفاده می کند).

۶. پیشگویی و درمان بیماران اورژانس

بخش اورژانس یک بیمارستان یکی از بخش های مهم بیمارستان بوده و معمولاً بیماران این بخش نیاز به مراقبت بسیار و تصمیمات سریع جهت درمان دارند. لین و همکاران [8] طی مطالعه ای در بیمارانی که در پایگاه داده بخش اورژانس یک بیمارستان تایوان به ثبت رسیده است سعی بر این داشتند تا به کشف بهترین تصمیمات جهت درمان پرداخته و بدین ترتیب از بسیاری از هزینه های ناشی از اقدامات نادرست نظیر جراحی بیهوشند. جهت انجام اینکار از روش خوشه بندی جهت دسته بندی مشخصه ها و علائم بیماری و تصمیمات اتخاذ شده توسط پرستار و پزشک با اجرای نرم افزار ROSE2 (Rough Sets Data Explorer) جهت استخراج داده کاوی پرداختند. بدین منظور از ترکیب الگوریتم k-means و الگوریتم نقشه خود سازمانده (SOM) از شبکه عصبی جهت انجام روش خوشه بندی استفاده گردید.

از نتایج حاصل از این مقاله می توان به موارد زیر اشاره کرد:

بیماران با زمان بستری بالا در اورژانس و مصرف بالای هزینه پزشکی و میانگین سن بالا، در گروه ریسک بالا قرار می گیرند و اغلب هزینه های بسیاری را مصرف می کنند و تشابهات زیادی با اغلب بیماران درجه ۱ و ۲ را دارند. در قسمت هزینه های بیمارستانی: بیماران بخش بیماریهای داخلی، بیشترین مصرف تجهیزات پزشکی را دارند و بیشتر بیماران توسط آمبولانس آورده می شوند.

۷. پیش بینی مدت زمان بستری بیماران

پیش بینی مدت زمان بستری شدن بیماران باعث مدیریت بهتر منابع بیمارستانی، ارائه خدمات بیشتر و افزایش رضایتمندی بیماران می گردد. در تحقیق انجام شده بر روی ۲۰۶۳۱۹۱ رکورد از پایگاه داده کودکان (زیر ۲۰ سال) بستری شده در ۳۰۰۰ بیمارستان ثبت شده در سال ۲۰۰۶، الگوریتم های داده کاوی برای ساخت مدلی به منظور پیش بینی مدت زمان بستری شدن بیماران دستگاه گوارش که نیاز به مراقب های کوتاه مدت دارند مورد تحلیل قرار گرفت. [5] به دلیل کاهش مقیاس بزرگ داده ها از نمونه های تصادفی از مجموعه رکوردها استفاده گردید و به جهت برآورد مقادیر مفقود شده، چهار سیاست اتخاذ شد. مقادیر پرت در ویژگی ها از چارک ها شناسایی شده و برای شناسایی رکوردهای پرت از الگوریتم تشخیص ناهنجاری، که از الگوریتم های خوشه بندی است، استفاده گردید. با توجه به اینکه بسیاری از ویژگی ها با اهداف تحقیق غیر مرتبط هستند، از ۱۶۲ ویژگی ۴۰ ویژگی که بر مدت زمان بستری شدن بیماران دستگاه گوارش تاثیر گذاری بیشتری داشتند شناسایی و مدل نهایی با استفاده از این ویژگی ها ساخته شد. از الگوریتم هایی مانند درخت تصمیم، شبکه های عصبی و الگوریتم های مبتنی بر بیز برای ساخت مدل با استفاده از داده های آموزشی استفاده شد. نتایج بدست آمده از دقت مدل های ایجاد شده بر روی داده های آزمایشی نشان می دهد که الگوریتم C.5 دقت بالاتری نسبت به سایر الگوریتم ها دارد. با این الگوریتم می توان با دقت ۸۲٪ مدت زمان بستری شدن بیماران دستگاه گوارش که نیاز به مراقبت های کوتاه مدت دارند را پیش بینی کرد.

۸. شناسایی و پیشگویی حملات قلبی

تحقیق بر روی ۳۰۳ داده در دو کلاس طبیعی (۱۶۴) و بیماران قلبی (۱۳۹) انجام شد. [6] جهت شناسایی و پیشگویی حملات قلبی از روشهای خوشه بندی داده کاوی استفاده گردید. خوشه بندی یکی از اصلی ترین اعمال داده کاوی است که هدف آن گروه بندی داده ها به کلاسهای معنی دار (خوشه ها) می باشد. به نحوی که شباهت بین داده های یک خوشه بیشترین و شباهت بین داده دو خوشه مجزا کمترین باشد. در این تحقیق با توجه به وجود داده های بیماران قلبی که در بسیاری از موارد ویژگی های آنها شامل اعداد مرکب یا طبقه بندی شده است، از ترکیب الگوریتم های ژنتیک و k-means استفاده گردید که نتایج خوبی را جهت شناسایی بهتر خوشه ها و در نتیجه تشخیص و پیشگویی حملات قلبی داشته است.

۹. کشف دانش پزشکی در انبار داده های بالینی طب سنتی

چینی

اخیراً طب سنتی چینی به عنوان مکمل های درمانی بسیاری از بیماریها نظیر سرطان، آرترویدروماتوئید، لوسمی و میگرن به تصویب رسیده است. [9] از آنجا که منبع اصلی کسب دانش تجربی این علم، شواهد بالینی طولانی مدت نظیر نتایج آزمایشگاهی، تشخیص و نسخه می باشد، به این دلیل انبار داده بسیار بزرگی را شامل می شود که تنها روشهای داده کاوی می توانند به کشف دانش در این حوزه بپردازند. در تحقیق انجام شده بر روی ۲۰۰۰۰ بیمار بستری طب سنتی چینی و ۲۰۰۰۰ بیمار سرپایی از روشهای خوشه بندی استفاده گردید که از میان ۵ روش خوشه بندی نظیر SVM، شبکه عصبی و درخت تصمیم، برای سندرهم های متفاوت در ۱۰۶۹ مورد اپیدمیولوژی بالینی، روش SVM نتایج پیشگوییانه بهتری را نشان داد. همچنین از قوانین وابستگی و تحلیل شبکه پیچیده (CAN) جهت یافتن نقاط مفید برای انجام عمل طب سوزنی و الگوهای ترکیب گیاهان دارویی، از تکنیک آموزش ماشین چندگانه (SVM) جهت پیشگویی شروع نفروپاتی و از شبکه های عصبی برای پیش بینی نقش اطلاعات تشخیصی در اثر درمان آرتوئید روماتوئید بهره جسته شد.

۱۰. نتایج

در این مقاله سعی شد تا یک بررسی جامع در خصوص هنر داده کاوی در حوزه های مختلف علوم پزشکی پرداخته گردد و یک راهنما و دستورالعمل جهت انجام مطالعات داده کاوی در این حوزه ارائه دهد. بدین منظور جهت سهولت انتخاب مناسب محققان، پزشکان، مهندسين و کارشناسان آمار از ابزارهای داده کاوی، نتایج حاصل از این تحقیق به صورت دسته بندی تطبیقی الگوریتم های داده کاوی در حوزه های مختلف علوم پزشکی در جدول ۱ آورده شده است.

۱۱. نتیجه گیری

با توجه به تحقیقات انجام شده مشخص گردید که داده کاوی در اکثر حوزه های علوم پزشکی به جهت پیشگویی بیماری از طریق شناسایی عوامل موثر بر بیماری، تشخیص و درمان بهینه کاربرد داشته است. همچنین می تواند در کاهش هزینه های بیمارستانی از طریق پیش بینی مدت زمان بستری، تشخیص نوع

- 2 Shahrabi J, Zolghadr Shojaei A. Advanced data mining: Concepts and algorithms. 2009; Tehran, Jahad daneshgahi.
- 3 Salalri M, Adibnia f. The top 10 algorithms for data mining. 13th Iranian Student Conference of Electronic Engineering. 2010.
- 4 Asadi N, Sadrodini M. Employing data mining to identify cancer risk factors and determine the optimal treatment in Namazi hospital cancer database. 16th Annual National Conference of Computer Society of Iran. 2010; Sharif University.
- 5 Oliyaei A, Salmasi N. An efficient model to predict duration of hospitalization of digestive system patients. 5th Conference of Data mining of Iran, 2011; Amirkabir University.
- 6 Dehghani T, Afshari Saleh M, Khalilzadeh M. A genetic K-means clustering algorithm for heart disease data. 5th Conference of Data Mining of Iran, 2011; Amirkabir University.
- 7 Pardalos PM, Tomaino V, Xanthopoulos P. Optimization and data mining in medicine. 2009; Springer.
- 8 Lin WT, Wu YC, Zheng JS, Chen MY. Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. Expert Systems with Applications, 2011; 38: 11078-11084.
- 9 Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Guo Y, Zhang H, Gao Z, Yan X. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artificial Intelligence in Medicine Journal, 2010; 48: 139-152.
- 10 Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 2008; 77: 81-97.

بیماری و اتخاذ بهترین تصمیم در درمان بیماران اورژانس تاثیرات به سزایی داشته است. روش های داده کاوی توانایی های زیادی را در کاوش داده های حجیم علوم آزمایشگاهی به جهت تحلیل داده های میکروسکوپی، بررسی رفتار ژنها و رشته های DNA نشان داده است. شیوه های پیشگویانه این علم نتایج بسیار مفیدی را نیز در طب سنتی به همراه داشته است که از آن میان می توان به ارائه بهترین الگوی ترکیب مواد گیاهی، شناسایی بهترین نقاط جهت عمل طب سوزنی، تشخیص عوامل بیماری زا، پیشگویی نفروپاتی دیابت و درمان آرتوئید روماتوئید اشاره نمود.

بررسی روشهای به کار رفته در حوزه های مختلف علوم پزشکی نشان داده است که از میان روشهای مختلف داده کاوی، بیشترین روش به کار رفته در این حوزه، روش خوشه بندی و دسته بندی بوده است که جهت انجام این کار از الگوریتم های بسیاری استفاده شده که از آن میان الگوریتم k-means نتایج پیشگویانه بهتری را نسبت به سایر الگوریتم ها داشته است.

References:

- 1 Shahrabi J, Shakoorniaz V. Concepts of data mining in Oracle 11. 2008; Tehran.

جدول ۱: دسته بندی روش ها و الگوریتم های داده کاوی به کار رفته در علوم پزشکی

ردیف	حوزه پزشکی مورد تحقیق	روش داده کاوی به کار رفته	الگوریتم داده کاوی به کار رفته	توضیحات
۱.	پیشگویی بیماری	خوشه بندی دسته بندی قوانین وابستگی	k-means apriori	تعیین عوامل موثر بر ایجاد انواع سرطان
۲.	تعیین بهترین نوع درمان	خوشه بندی دسته بندی قوانین وابستگی	k-means apriori	تعیین بهترین نوع درمان سرطان از میان ۳ روش: جراحی - شیمی درمانی - رادیوتراپی
۳.	بررسی داده های آزمایشگاهی	خوشه بندی	SAMBA	بررسی رفتار ژن ها با رشته های DNA جهت پیشگویی اختلالات ژنتیکی و بیماریهای مهلک
۴.	پیشگویی در بیماران اورژانس	خوشه بندی	ترکیب k-means و الگوریتم نقشه خود سازمانده (SOM) از شبکه عصبی	جهت اتخاذ تصمیم درمانی صحیح و به موقع و کاهش هزینه های بیمارستانی
۵.	پیش بینی مدت زمان بستری	خوشه بندی دسته بندی	درخت تصمیم شبکه های عصبی Naïve Bays C 4.5	پیش بینی مدت زمان بستری شدن بیماران دستگاه گوارش که نیاز به مراقب های کوتاه مدت دارند جهت کاهش هزینه های بیمارستانی
۶.	شناسایی و پیشگویی عارضه های بیماری	خوشه بندی	ترکیب الگوریتم ژنتیکی و الگوریتم K-means	شناسایی و پیشگویی حملات قلبی
۷.	طب سنتی چینی	خوشه بندی	ماشین بردار پشتیبان (SVM) درخت تصمیم شبکه Bayesian	کشف سندروم های مختلف
۸.	طب سنتی چینی	قوانین وابستگی	تحلیل شبکه پیچیده CNA	یافتن نقاط مطلوب طب سوزنی و الگوهای ترکیب گیاهان دارویی
۹.	طب سنتی چینی	خوشه بندی	ماشین بردار پشتیبان (SVM)	پیشگویی شروع نفروپاتی دیابت
۱۰.	طب سنتی چینی	خوشه بندی	شبکه های عصبی	درمان آرتريت روماتوئید