

امکان سنجی اجرای تکنیک های داده کاوی در بانک های اطلاعاتی در حوزه بهداشت و درمان سازمان تامین اجتماعی

مهکامه مهدی پور^۱، بابک درویش روحانی^۲

^۱ دانشجوی کارشناسی ارشد نرم افزار، دانشگاه پیام نور هرمزگان مرکز بین المللی قشم.

^۲ عضو هیئت علمی دانشگاه پیام نور هرمزگان مرکز بین المللی قشم.

نام نویسنده مسئول:

مهکامه مهدی پور

تاریخ دریافت: ۱۴۰۱/۱۱/۱۶

تاریخ پذیرش: ۱۴۰۲/۰۱/۲۱

چکیده

با استفاده از کلان داده در زمینه بهداشت و درمان، می توان خدمات بهتری را به مردم ارائه داد. کاربرد کلان داده در پزشکی شناسایی روش های شخصی سازی شده برای بیماران است. همین شخصی سازی است باعث افزایش سلامت در جامعه خواهد شد و همچنین هزینه های دولت نیز در بخش بهداشت و درمان کاهش خواهد یافت. تعداد و اندازه پایگاه داده های پزشکی به سرعت در حال افزایش است و مدل های توسعه یافته تکنیک داده کاوی میتوانند برای پزشکان جهت کمک در تصمیم گیری موثر و کاربردی باشند. هدف اصلی از این مقاله، گزارش یک پروژه تحقیقاتی به منظور مقایسه الگوریتم های مختلف داده کاوی از طریق مقایسه حساسیت، ویژگی و دقت بین آنها، جهت انتخاب دقیق ترین مدل برای پیش بینی عود مجدد انواع بیماری ها بوده است. این تحقیق به بررسی تحقیقات انجام شده در زمینه بهره برداری از داده کاوی در حوزه بهداشت و درمان به خصوص در سازمان تامین اجتماعی پرداخته است. نتایج حاصل از این تحقیق می تواند کمکی برای سازمان تامین اجتماعی در راستای یافتن قدم های آینده در جهت نیل به اهداف سازمان باشد. باتوجه به جمعیت روزافزون بیمه شدگان این سازمان بهره گیری از تکنیک های داده کاوی میتواند به سازمان تامین اجتماعی در جهت دسترسی سریع تر و راحت تر به اطلاعات سازمان و اجرای راحت تر سیاست های جدید و دسته بندی منظم و مشخص جهت ارتقا ارتباطات و داشتن کارایی برای ارتباط مؤثر و بیشتر کمک کند.

واژگان کلیدی: داده کاوی، بیگ دیتا، تامین اجتماعی، بانک اطلاعاتی، کلان داده.

مقدمه

روزانه حجم عظیمی از داده‌ها توسط سیستم‌های اطلاعاتی (*information systems*) مدرن، فناوری‌های دیجیتال مانند اینترنت اشیا (*Internet of Things*)، رایانش ابری (*cloud computing*) و دیگر موارد تولید می‌شود (۱). تحلیل این داده‌های انبوه که به آن‌ها کلان داده (*Big Data*) گفته می‌شود نیازمند تلاش‌های زیاد در سطوح گوناگون، جهت استخراج دانش به منظور کمک به تصمیم‌سازی بهتر است. (۲) بنابراین، تحلیل کلان داده یک حوزه علمی و صنعتی روز محسوب می‌شود که پژوهش‌های زیادی پیرامون آن در جریان است و همچنان نیازمند تحقیقات بیشتر نیز هست (۲).

کلان داده مجموعه داده‌های پیچیده و بزرگی گفته می‌شود که باید پردازش و تجزیه و تحلیل شوند تا اطلاعات ارزشمندی که می‌تواند به سود کسب و کارها و سازمان‌ها باشد، کشف شوند (۳). تنوع کلان داده‌ها به داده‌های ساختاریافته، بدون ساختار و نیمه ساختاریافته گفته می‌شود که از چندین منبع جمع‌آوری شده‌اند (۴). در حالی که در گذشته داده‌ها فقط از صفحه گسترده و پایگاه داده جمع‌آوری می‌شدند، امروزه به صورت آرایه‌ای از جمله ایمیل‌ها، PDFها، عکس‌ها، فیلم‌ها، فایل‌های دیداری، پیام کوتاه و موارد دیگر ارائه می‌شوند. تنوع یکی از خصوصیات مهم داده‌های کلان است (۱).

حجم یکی از ویژگی‌های داده کلان است. ما می‌دانیم که داده حجیم، حجم زیادی از داده را نشان می‌دهد که به طور روزانه از منابع مختلفی مانند رسانه‌های اجتماعی، فرایندهای کسب و کار، ماشین‌آلات، شبکه‌ها، تعاملات انسانی و غیره تولید می‌شود. این حجم زیاد از داده‌ها در انبارهای داده ذخیره می‌شوند (۵). با بینش کلان داده، همیشه می‌توانید یک قدم جلوتر از رقبا باشید. شما می‌توانید بازار را بشناسید تا بدانید رقبای شما چه نوع تبلیغات و پیشنهادهایی را ارائه می‌دهند و سپس می‌توانید پیشنهادهای بهتری برای مشتریان خود ارائه دهید (۴). همچنین، بینش کلان داده به شما امکان می‌دهد تا رفتار مشتری را یاد بگیرید تا روند او را درک کنید و تجربه‌ای کاملاً "شخصی" را به او ارائه دهید تجزیه و تحلیل کلان داده می‌تواند به شرکت‌ها کمک کند تا منجر به فروش بیشتری شوند که طبیعتاً به معنای افزایش درآمد است (۵).

کسب و کارها از ابزارهای تجزیه و تحلیل کلان داده استفاده می‌کنند تا درک کنند که محصولات و خدمات آن‌ها در بازار چگونه عمل می‌کنند و مشتریان چگونه به آن‌ها پاسخ می‌دهند (۶و۷). بنابراین، بهتر می‌توان درک کرد که کجا وقت و پول خود را سرمایه‌گذاری می‌کنند (۸).

ترکیب نتایج و روش‌ها:

کلید واژه‌های داده کاوی، تامین اجتماعی، کلان داده، دیتا ماینینگ در پایگاه داده‌های اسپرینگر، الزویر جستجو شده اند نتایج به شرح زیر است:

کلان داده در حال حاضر ایجاد اختلاف فاحشی در بخش مراقبت‌های بهداشتی را آغاز کرده است (۹). با کمک تجزیه و تحلیل‌های پیش‌بینی شده، متخصصان پزشکی و *HCP* اکنون می‌توانند خدمات مراقبت‌های بهداشتی شخصی را برای بیماران ارائه دهند (۱۰). کنترل و دستیابی به بیماران خیلی سریع انجام می‌گیرد. جدا از این، تناسب اندام، پزشکی از راه دور، نظارت از راه دور؛ همه با استفاده از کلان داده و هوش مصنوعی به تغییر زندگی به سمت بهتر کمک می‌کنند. با کمک تحلیل دقیق داده‌ها می‌توانید به اطلاعات تعداد زیادی از بیماران در سنین مختلف و گروه‌های متنوع برسید و برای مقابله با بیماری‌ها و پیشگیری از ابتلای به آنها، راهکارهای مناسب و کاربردی ارائه دهید (۱۱).

با استفاده از کلان داده در زمینه بهداشت و درمان، می‌توان خدمات بهتری را به مردم ارائه داد. کاربرد کلان داده در پزشکی شناسایی روش‌های شخصی سازی شده برای بیماران است. همین شخصی سازی است باعث افزایش سلامت در جامعه خواهد شد و همچنین هزینه‌های دولت نیز در بخش بهداشت و درمان کاهش خواهد یافت (۱۲). می‌توان از سوابق بیماران به خصوص مربوط به داروهای داده محور استفاده کرد و با یک تحلیل گسترده، الگوهایی را برای کمک به دانشمندان جهت شناخت بیماری‌ها، درمان آنها و تولید داروهای جدید ارائه داد. چون بیگ دیتا‌ها به طور بالقوه می‌تواند اطلاعات بیشتری را در مورد اثربخشی داروها و درمان‌های داده شده یک فرد خاص با مجموعه‌ای از علائم و بیماری‌های خاص را نشان دهد. مجموعه کلان داده‌ها با نظارت دقیق بر مشکلات پزشکی بیماران و با استفاده از اطلاعات ژنتیکی بدست آمده در نهایت قادر

به فاش کردن اطلاعات خاص در مورد بیمار خواهد شد که ممکن است بدون استفاده از آنها، این اطلاعات به سادگی بدست نیابند (۱۳)

صنعت بهداشت و درمان به سرعت درگیر تغییرات فناوری است. یک تغییر تدریجی از الگوی خدمات بهداشتی مبتنی بر خدمات که در درجه اول بر روی درمان بیماریها متمرکز شده اند (۱۴)، به یک رویکرد نتیجه گرایانه که نه تنها شامل بررسی روش های مختلف درمانی است بلکه در واقع با هدف درک عوامل ایجاد کننده بیماری های مختلف و از بین بردن آنها نیز می باشد (۱۴ و ۱۵). افزایش قابل توجهی در داده های بهداشتی، چه ساختاری و چه غیر ساختاری تولید نشده وجود دارد (۱۶). سطح پیچیدگی بالای این داده ها مستلزم آن است که توسط تجزیه و تحلیل داده های بزرگ پردازش شده، استنتاج های مربوط به آن بررسی شود و قابل اجرا ارائه شود (۱۷). *Telemedicine* همچنین در حال بهره برداری از قدرت داده های بزرگ برای بهبود امکانات مراقبت های بهداشتی موجود است (۱۷).

قدرت پیش بینی داده ها می تواند یک دارایی بسیار ارزشمند باشد، اما دستیابی به بینش از کلان داده (*Big Data*) نیاز به مجموعه مهارت های ظریفی دارد. یکی از این مهارت ها داده کاوی است (۱۴). داده کاوی با کاوش در داده ها به شما امکان می دهد الگوهای مهمی را که می تواند در کار تجزیه و تحلیل داده ها و تصمیمات مربوط به کسب و کار در آینده کمک کند، مشاهده نماید (۱۵).

داده کاوی

داده کاوی قادر به کشف و استخراج دانش جدید از داده های گذشته نگر است. نحوه پیش پردازش داده ها و هم چنین متغیرهای منتخب، تاثیر قابل توجهی در کشف دانش دارد (۱۴).

داده کاوی در واقع یک روش است، که به واسطه آن مسئله ای حل شود. این روش بر روی حجم زیادی از داده ها کار می کند و تحلیل مورد نیاز را انجام می دهد، در نهایت یک سری الگوهای تکرار شونده استخراج می شود که می تواند برگ برنده باشد. سپس نوبت به پیدا کردن ارتباطات بین الگوهای مختلف است و در نهایت یک سری راه حل های مهم در این چالش وارد می شوند (۱۰).

شرکت های بزرگی که از داده کاوی استفاده می کنند سعی می کنند این رویکرد را برای تحلیل رقبا و بازار در نظر بگیرند و به واسطه این روش پیش بینی خوبی بر روی ترندهای روز خواهند داشت (۱۱). از سوی دیگر برنامه هایی که توسط این شرکت ها برای آینده در نظر گرفته می شود هم جهت با خواسته عموم افراد است و این روال سبب می شود موقعیت پیشی گرفتن از رقبا آسانتر شود. فرایند داده کاوی بخش های مختلفی دارد و شاید صحبت در مورد آن زمان و حوصله زیادی نیاز داشته باشد اما موضوع از این قرار است که این علم به طور خلاصه سه بخش مهم را در بر می گیرد (۱۲). بخش اول: استخراج داده ها از بخش های مختلف و انتقال آنها در پایگاه های داده ای که قابلیت چند بعدی داشته باشند و در ذخیره سازی اطلاعات مهمی که در آینده مورد استفاده قرار می گیرد. بخش دوم: انتقال داده ها به لایه های مختلف از جمله لایه های کسب و کار که بر اساس نرم افزار داده کاوی دنبال می شود. بخش سوم: تحلیل داده ها، نمایش نتایج حاصل از تحلیل داده که اغلب به صورت فرم ساده یا در گراف و نمودار ارزیابی می شود. منظور از دیتا در علم داده کاوی می تواند یک سری داده هایی باشد که توسط مردم و در ارتباطات روزمره کسب شده است، می تواند شامل یک سری پیش بینی ها و احتمالات باشد و یا داده های منطقی باشد که در پایگاه داده ها به ذخیره سازی رسیده است.

هر داده ای که دریافت شود نیاز به پیش پردازش و پس پردازش دارد و این قدم بسیار مهم است (۴). مشخص است که در نهایت باید داده ها به وضعیت پیاده سازی برسند و برای پیاده سازی می بایست الگوریتم مناسبی استفاده شود که موقعیت داده کاوی را به خوبی می سنجد (۵). الگوریتم هایی که در این راستا استفاده می شوند اغلب شامل کلاس بندی، خوشه بندی و یا یک سری الگوهای یادگیری تقویتی هستند که می توانند ارتباطی که بین داده ها وجود دارد، را به طور دقیق ارزیابی کنند (۴ و ۵). هر تکنیکی که برای داده کاوی استفاده می شود سبب می شود تا سرعت انجام محاسبات بیشتر شود و از سوی دیگر فضای مورد نیاز برای حافظه نیز بهبود داشته باشد. در ادامه سه تکنیک مهم در راستای داده کاوی را بررسی می کنیم (۴).

تکنیک طبقه بندی: طبقه بندی که به آن *classification* گفته می‌شود، یکی از روش‌های مهم داده کاوی محسوب می‌شود. این الگوریتم از روش برچسب زنی داده‌ها استفاده می‌کند. به طوری که هر داده بر اساس ویژگی که برای آن تعریف شده است برچسب گذاری می‌شود و همینطور در کلاس‌های مختلفی قرار می‌گیرد. این الگوریتم‌ها خودآموز هستند یعنی روش‌های جدید برچسب گذاری را یاد می‌گیرند و طبق همین ویژگی می‌توانند یک سری نمونه‌های جدید را برچسب بزنند. این تکنیک از داده کاوی مبتنی بر یادگیری است که یک سری مدل‌های خوب بر روی داده‌های جدید را اعمال می‌کند و پس از آن گروه‌بندی مشتریان به درستی صورت می‌گیرد (۴).

تکنیک خوشه‌بندی: خوشه‌بندی یا کلاستر از دیگر تکنیک‌هایی است که در علم داده کاوی اهمیت زیادی دارد. این روش به صورتی است که ذات داده اهمیت زیادی در آن دارد و گروه‌بندی بر اساس ذات داده‌ها صورت می‌گیرد. هر داده‌ای که در یک گروه قرار می‌گیرد با ذات اصلی همه اعضای گروه همخوانی دارد. این تکنیک فروشگاه‌های مختلف اجرا می‌شود، مثلاً یک گروه شامل مشتریانی هستند که علاقه به خرید با قیمت کم اما دفعات زیاد دارند و گروهی دیگر نیز افرادی هستند که کم خرید می‌کنند اما بهترین‌ها را انتخاب می‌کنند (۴).

تکنیک یادگیری تقویتی: این روش به صورتی دنبال می‌شود که الگوریتم سعی می‌کند اطلاعات و عملیاتی با محیط پیرامون ایجاد نماید و سعی به کشف اطلاعات داشته باشد و به همینطور منظور یادگیری خود را نیز پیوسته‌تر دنبال کند. در این روش شبیه‌سازی بسیار دیده می‌شود (۴).

به طور کلی، داده کاوی شامل شش وظیفه اصلی است: (۵) تشخیص ناهنجاری شامل شناسایی انحراف در یک مجموعه داده است. بسته به زمینه، این انحرافا ممکن است نشان‌دهنده خطاهای داده یا پرتوهای اطلاعاتی باشند (۵).

یادگیری قاعده ارتباط، یک روش یادگیری ماشین است که برای شناسایی همبستگی‌های مفید بین متغیرها استفاده می‌شود. به عنوان مثال، بانک‌ها از این روش برای شناسایی محصولات که مشتریان معمولاً با هم خریداری می‌کنند، استفاده کرده و آن را در تدوین استراتژی‌های فروش به کار می‌گیرند (۵).

خوشه‌بندی، شناسایی گروه‌هایی از سوابق یا ساختارها در یک مجموعه داده است که دارای یک ویژگی مشترک هستند، به عنوان مثال گروه‌بندی بر اساس رنگ مو (۵).

طبقه‌بندی شامل استفاده از آنچه قبلاً در مورد یک مجموعه داده می‌دانید برای دسته‌بندی داده‌های جدید (به عنوان مثال طبقه‌بندی مشتریان بر اساس محدوده سنی و موقعیت مکانی آن‌ها) است (۵).

تحلیل رگرسیون روابط بین یک یا چند مقدار را برجسته می‌کند. به طور خاص، متغیرهای مستقل چگونه بر متغیرهای وابسته تأثیر می‌گذارد؟ (به عنوان مثال تأثیر سن یا رژیم غذایی بر وزن کسی) (۴).

خلاصه‌برداری اوج تمام مراحل است که توضیح دادیم. این مرحله شامل ایجاد یک گزارش روشن و مختصر از یافته‌های شما، معمولاً با تجسم است (۴).

داده کاوی اساساً یک عالم کوچک از کل فرآیند تجزیه و تحلیل داده است (۱). در واقع، درجه بالایی از هم‌پوشانی میان این دو وجود دارد (۱). تفاوت اصلی در ابزارها و تکنیک‌ها نیست، بلکه در دلایلی است که شما از آن‌ها استفاده می‌کنید (۲). در حالی که تجزیه و تحلیل داده شامل آزمایش فرضیه‌ها است، داده کاوی با استفاده از همان روش‌ها، الگوها را در کلان داده شناسایی می‌کند. بر اساس این الگوها، می‌توانید فرضیه بعدی خود را شکل دهید (۳).

نیاز به امنیت در ابعاد مختلف اجتماعی و اقتصادی، یکی از پایه‌ای‌ترین نیازها در جوامع انسانی است. حمایت‌هایی که امروزه در مجموع از آنها تحت عنوان "تأمین اجتماعی" یاد می‌شود، مفهومی تازه و متعلق به عصر جدید نیست و پیدایش آن را می‌توان به شکل‌گیری نخستین اجتماعات انسانی و جوامع مدنی و همچنین شکل‌گیری و تأسیس دولت‌ها در جوامع مرتبط دانست (۲).

همانگونه که در قانون ساختار نظام جامع رفاه و تأمین اجتماعی - مصوب سال ۱۳۸۳ مجلس شورای اسلامی - اشاره شده، سه قلمرو متمایز "حمایتی"، "امدادی" و "بیمه‌ای" برای نظام جامع تأمین اجتماعی کشور تعریف شده که سازوکارها و نحوه تأمین منابع در این سه قلمرو، کاملاً متمایز و متناسب با شرح وظایف و ماموریت‌های تعریف شده برای هر قلمرو است (۳).

آنچه در این بین مسلم میزان جامع و گستردگی سازمان است که به دنبال آن حجم بسیار عظیمی از دیتا را خواهد داشت(۲).

با توجه به گسترش روز افزون جوامع و نیاز به تامین اجتماعی بیشتر احساس میشود و از طرفی صنعت بهداشت و درمان به سرعت درگیر تغییرات فناوری است. یک تغییر تدریجی از الگوی خدمات بهداشتی مبتنی بر خدمات که در درجه اول بر روی درمان بیماریها متمرکز شده اند، به یک رویکرد نتیجه گرایانه که نه تنها شامل بررسی روش های مختلف درمانی است بلکه در واقع با هدف درک عوامل ایجاد کننده بیماری های مختلف و از بین بردن آنها نیز میباشد. افزایش قابل توجهی در داده های بهداشتی، چه ساختاری و چه غیر ساختاری تولید نشده وجود دارد. سطح پیچیدگی بالای این داده ها مستلزم آن است که توسط تجزیه و تحلیل داده های بزرگ پردازش شده، استنتاج های مربوط به آن بررسی شود و قابل اجرا ارائه شود(۴).

جهت پردازش این داده ها و دسترسی اسان تر و کنترل و پیاده سازی تصمیمات کلان سازمانی نیازمند الگو و روشی هستیم که مشکلات مطرح شده را به حداقل برساند(۱۱).

نتایج

نتایج حاصل از این تحقیق می تواند کمکی برای سازمان تأمین اجتماعی در راستای یافتن قدم های آینده در جهت نیل به اهداف سازمان باشد. همچنین می تواند به سازمان تأمین اجتماعی در جهت دسترسی سریع تر و راحت تر به اطلاعات سازمان و اجرای راحت تر سیاست های جدید و دسته بندی منظم و مشخص جهت ارتقا ارتباطات و داشتن کارایی برای ارتباط مؤثر و بیشتر کمک کند. تغییر و تبیین راحت تر قوانین جدید بذای جامعه بزرگ. بررسی عملکرد های تعهدات علی الخصوص تعهدات بین نسلی. دسته بندی گروه های خدمت گیرنده و ارزیابی آنها در صورت تعیین یک الگوریتم کارآمد..

منابع و مراجع

- [۱] پایان نامه کاربرد داده کاوی در تحلیل داده های سازمان تامین اجتماعی، "۱۳۸۵،
- [۲] پایان نامه تحلیل و شناسایی مراجعین مراکز درمانی سازمان تامین اجتماعی با استفاده از روش های داده کاوی، ۱۳۹۱، محسن امامی.
- [۳] داده کاوی اطلاعات در مدیریت زنجیره تامین جهت افزایش رضایت مندی مشتریان مزگان جعفرزاد، دانشگاه آزاد اسلامی واحد تهران جنوبه.
- [۴] طلوعی اشلقی عباس، پورابراهیمی علی، ابراهیمی ماندانا، قاسم احمد لیلا. پیش بینی عود مجدد سرطان پستان به کمک سه تکنیک داده کاوی.
- [۵] خورسندغفاری، توسلی، سوادی. کاربردهای متفاوت داده کاوی. In دومین کنفرانس ملی توسعه علوم مهندسی May 13 ۲۰۱۵.
- [۶] فرد، محمد تقی. داده کاوی و کاربرد آن بیش بینی الگوهای سلامت. فصلنامه دانش مدیریت (منتشر نمی شود). ۲۰۰۹ (Feb 20;21).
- [۷] فرد، محمد تقی تقوی، ناصرزاده، سید محمد رضا، فراست. داده کاوی و کاربرد آن در تصمیم گیری ها. فصلنامه دانش مدیریت (۲۰۰۸) May 21;21 (۲).
- [۸] مقدسی حمید، حسینی اعظم السادات، اسدی فرخنده، جهانبخش مریم. داده کاوی و کاربرد آن در سلامت.
- [۹] توکلی، احمد، مرتضوی، کاهانی، حسینی. به کارگیری فرایند داده کاوی برای پیش بینی الگوهای رویگردانی مشتری در بیمه. چشم انداز مدیریت بازرگانی. ۲۰۱۱ Feb 20;9.
- [۱۰] مرتضوی، کاهانی. داده کاوی برای پیش بینی الگوهای تمایل مشتری در بیمه. چشم انداز مدیریت بازرگانی. Feb 21;9 ۲۰۱۳.
- [۱۱] Boskabadi M, Doostparast M. مدل بندی و داده کاوی داده های جهانی بیماران ویروس کووید ۱۹. Iranian Journal of Emergency Medicine. 2020 Oct 12;7(1):e40.
- [۱۲] کاراموزیان، خزاعی، خسروی راد. نقش داده کاوی و مدیریت دانش در مدیریت ارتباط با مشتری. پژوهش های جدید در مدیریت و حسابداری. ۲۰۲۱ Sep 23;76(7):173-84.
- [۱۳] ضیایی، ناجی پژمان، نقی زاده، مختاباد. بررسی امکان مدل سازی میزان تغییرات کالبدی بافت های شهر با استفاده از داده کاوی. هویت شهر. ۲۰۲۲ Oct 23;16(3):5-16.
- [۱۴] دکامینی فاطمه، احسانی فر محمد. مقایسه ی کارایی الگوریتم های داده کاوی در پیش بینی تشخیص بیماری دیابت. ۲۰۲۱.
- [15] Calle J. Breast cancer facts and figures 2003-2004. American Cancer Society 2004. <http://www.cancer.org/>. (last accessed: Jan. 2010).
- [16] Karabatak M, Cevdet M. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 2009; 36: 3465–9.
- [17] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques, 2nd Edition. San Fransisco: Morgan Kaufmann; 2005.