7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

Investigating the Challenges of Arranging a Data Mining Service in the Field of Health Information Systems

Mohammad Hassan Zarezade[1], Raouf Khayami[2]

[1] Master student, Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz
mammad1845@gmail.com
[2] Assistant professor, Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz
khayami@sutech.ac.ir

Abstract

The knowledge of data mining is one of the developing and superior knowledge in the world. This knowledge has been used in various fields such as commerce, electricity, industry and etc. one of the most essential usage of this knowledge is in the health sector that nowadays it has paid special attention to it. In this research we tried to review the following challenges of data mining in health and to accomplish a complete and exhaustive study in this field.at the first, in this research, the literature on the subject and the basic concepts are discussed. Undeniably, any new knowledge will have drawbacks and issues in the beginning and will also produce challenges that will lead the way for further research in this case and also, help to advance this field to enhance and complete it in the future, which will point to several important and vital challenges ahead for the conduct of data mining techniques in the health information system (HIS) including data quality, holding privacy and etc. and then, some of completed works on the challenges and issues will be presented.

Keywords: Data Mining, Health Information System, Data Quality, Privacy

7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

بررسی چالش‌های استقرار سرویس داده کاوی در حوزه سیستم‌های اطلاعات سلامت

*محمد حسن زارع زاده[1]، رئوف خیامی[2]

[1] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز،
mammad1845@gmail.com

[2] استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز
khayami@sutech.ac.ir

چکیده

دانش داده کاوی یکی از دانش‌های درحال توسعه و برتر جهان است. این دانش در حوزه‌های مختلفی هم چون تجارت، برق، صنایع و... نیز به کار گرفته شده است. یکی از مهم‌ترین کاربردهای این دانش دربخش بهداشت و درمان نیز می‌باشد که امروزه توجه ویژه‌ای به آن شده است. در این تحقیق سعی شده که مروری بر چالش‌های پیش رو داده کاوی در بهداشت و درمان صورت پذیرد و بررسی کامل و جامعی در این زمینه انجام گیرد. در ابتدا در این تحقیق به ادبیات موضوع و مفاهیم اولیه پرداخته می‌شود. بدیهی است هر دانش نوین و نوپایی در ابتدای امر معایب و مشکلاتی به همراه خواهد داشت و چالش‌هایی را نیز به وجود می‌آورد که خود زمینه ساز تحقیقات بیشتر در این مورد خواهد شد و همچنین کمک به پیشرفت این حوزه جهت بهبود و تکمیل آن در آینده نیز خواهد کرد، که اشاره به چندین چالش مهم و حیاتی پیش رو برای اجرای تکنیک‌های داده کاوی در سیستم اطلاعات سلامت (HIS) از جمله کیفیت داده‌ها، حفظ حریم خصوصی و ... خواهد شد. و سپس برخی از کارهای انجام شده در زمینه چالش‌ها و مشکلات مطرح شده ارائه خواهد شد.

کلمات کلیدی

داده کاوی، سیستم اطلاعات سلامت، کیفیت داده‌ها، حفظ حریم خصوصی

۱- مقدمه

سازمان‌های مراقبت بهداشتی که با مشکل حجم زیاد داده‌ها برای ذخیره‌سازی و هزینه‌های سنگین نگهداری رو به رو بودند، امروزه با استفاده از سیستم‌های اطلاعات سلامت[1] (HIS) که حجم عظیمی از داده‌ها اعم از داده‌های بیماران، پزشکان و ... در آن ذخیره می‌شود، برطرف گردیده است. در این سیستم، داده‌ها با روش‌هایی همچون سوابق بیمه، نظر سنجی توسط بیماران، پرونده‌های الکترونیک سلامت[2] (EHR) و ... جمع‌آوری می‌شود. داده‌های جمع‌آوری شده حاوی اطلاعات ارزشمندی است که با تجزیه و تحلیل مناسب می‌توان آن‌ها را استخراج نمود. همچنین تجزیه و تحلیل داده‌های بهداشتی با افزایش عملکرد وظایف مربوط به مدیریت بیمار، مراقبت‌های بهداشتی بیماران را بهبود می‌بخشد[1].

تمامی مؤسسات مراقبت‌های بهداشتی نیاز به تجزیه و تحلیل تخصصی داده‌های پزشکی خود دارند، که این کار زمانبر و پرهزینه است. توانایی استفاده از داده‌ها در پایگاه‌های داده به منظور استخراج اطلاعات مفید برای مراقبت‌های بهداشتی با کیفیت، رمز موفقیت مؤسسات مراقبت‌های بهداشتی است[2]. در این زمان صحبت از داده کاوی به میان خواهد آمد که می‌تواند حجم گسترده‌ای از داده‌ها را مورد تحلیل قرار دهد. برای بدست آوردن نتایج مطلوب بایستی دانش و اطلاعات کافی از الگوریتم‌های مختلف داده کاوی را دارا بود. از جمله این روش‌ها و تکنیک‌ها که در سیستم‌های HIS به کار گرفته شده می‌توان به: خوشه‌بندی، طبقه‌بندی، درخت تصمیم، K – نزدیک‌ترین همسایه و ... اشاره کرد.

باید توجه داشت که داده کاوی در بهداشت یک حرکت نوپا است و به همین دلیل با موانع و چالش‌های متعددی رو به رو شده و برای عملکرد بهینه آن در حوزه بهداشت و درمان بایستی به این موانع توجه ویژه‌ای شود تا بتوان با صرف وقت و هزینه کمتری به نتایج مطلوبی رسید.

۲- مروری بر ادبیات موضوع

۲-۱-تعریف HIS

امروزه، استفاده گسترده از فن‌آوری‌های اطلاعات و ارتباطات[۳] (ICT) تقریباً در همه جنبه‌های زندگی از جمله‌بخش مراقبت‌های بهداشتی نفوذ کرده است. HIS برای استفاده کامل به خصوص از اینترنت در ارائه خدمات بهداشتی بهتر معرفی شد. HIS اغلب به تعامل بین مردم، فرآیند و فناوری برای پشتیبانی از عملیات و مدیریت در ارائه اطلاعات ضروری به منظور بهبود کیفیت خدمات بهداشتی و درمانی اشاره دارد. هدف HIS بهبود کیفیت مراقبت از بیمار است[3].

HIS زمینه‌های تصمیم‌گیری را فراهم می‌کند و دارای چهار عملکرد اصلی است: تولید داده، گردآوری، تجزیه و تحلیل، سنتز، برقراری ارتباط و استفاده. HIS، داده‌ها را ازبخش بهداشت و سایربخش های مربوط جمع‌آوری می‌کند، سپس داده‌های جمع‌آوری شده را تجزیه و تحلیل کرده و پس از اطمینان از کیفیت، ارتباط و به موقع بودن آنها، داده‌ها جهت تصمیم‌گیری در ارتباط با سلامت، به اطلاعات تبدیل می‌کند[4]. HIS به طور گسترده در سازمان‌های مراقبت‌های بهداشتی مورد استفاده قرار می‌گیرد تا از پردازش داده‌های متداول از جمله صورتحساب بیمار، حسابداری، کنترل موجودی، محاسبه آمار و نگهداری سابقه بیمار پشتیبانی کند. [5].

۲-۲-چالش HIS

در سال ۱۸۵۴ جان اسنو که پدر اپیدمیولوژی مدرن محسوب می‌شود، با استفاده از نمودارهایی توانست مکان انتقال وبا را که از طریق منبع آب بود پیدا کند. اسنو تعداد مرگ‌های ناشی از وبا را شمرد و آدرس قربانیان را روی نقشه لندن به عنوان میله‌های سیاه طراحی کرد. وی کشف کرد که بیشتر مرگ‌ها به سمت یک پمپ آب خاص در لندن جمع شده است. اسنو قادر بود شخصاً داده‌ها را در زمان خود جمع‌آوری و تجزیه و تحلیل کند زیرا حجم اطلاعات قابل کنترل بود. امروزه با توجه به اندازه جمعیت، حجم داده‌ها و سرعت شیوع بیماری، تحقق کاری که اسنو انجام داده تقریباً غیرممکن است.

اینجاست که داده کاوی برای مراقبت‌های بهداشتی مفید می‌شود ... [6].

۲-۳- داده کاوی

اندازه داده‌ها به طور کلی روز به روز در حال رشد است. اکنون نیاز به درک مجموعه داده‌های بزرگ، پیچیده و غنی شده با اطلاعات در زمینه‌های مختلف از جمله فناوری، تجارت و علوم افزایش یافته است[7]. برای ذخیره چنین حجم زیادی از داده‌ها یا اطلاعات، اندازه پایگاه‌های داده به سرعت افزایش می‌یابد. این پایگاه‌ها اطلاعات بسیار مفیدی دارند که ممکن است برای فرایند تصمیم‌گیری در هر زمینه‌ای بسیار مفید باشد. این کار با کمک داده کاوی یا دانش در پایگاه داده[۴] (KDD) امکان پذیر است[8]. تعاریف مختلفی از اصطلاح داده کاوی وجود دارد. یکی از پر کاربردترین تعاریف بیان می‌کند که "داده کاوی تجزیه و تحلیل مجموعه داده‌های مشاهده‌ای (غالباً بزرگ) برای یافتن روابط نامشخص و جمع‌بندی داده‌ها به روش‌های جدید است که برای صاحب داده قابل فهم و مفید باشد"[9].

۲-٤-اهداف داده کاوی در HIS

داده کاوی به طور عمده الگوهای معنی داری را که قابل شناسایی نبودند، استخراج می‌کند. سپس این الگوها می‌توانند با دانش ادغام شده و به کمک این دانش تصمیمات اساسی گرفته شود.

همچنین تکنیک‌های داده کاوی در حوزه بهداشت و درمان بسیار مفید هستند. آنها خدمات پزشکی بهتری را به بیماران ارائه می‌دهند و در تصمیمات مختلف مدیریت پزشکی به سازمانهای بهداشتی کمک می‌کنند. برخی از خدمات ارائه شده توسط تکنیک‌های داده کاوی در مراقبت‌های بهداشتی عبارتند از: تعداد روزهای اقامت در بیمارستان، رتبه‌بندی بیمارستان‌ها، درمان‌های موثرتر، ادعاهای بیمه کلاهبرداری توسط بیماران و همچنین ارائه دهندگان، بستری مجدد بیماران، شناسایی روش‌های درمانی بهتر برای گروه خاصی از بیماران، ساخت سیستم‌های توصیه داروی مؤثر و ... . با توجه به همه این دلایل، محققان بسیار تحت تأثیر قابلیت‌های داده کاوی قرار دارند[8].

۳- چالش‌های داده کاوی در HIS

۳-۱مقدمه

استفاده از داده کاوی در زمینه پزشکی به دلیل ویژگی‌های خاص حرفه پزشکی اقدامی بسیار چالش برانگیز است[6]. برخی تحقیق‌ها در زمینه چالش‌های استفاده از تکنیک‌های داده کاوی در HIS صورت گرفته است. همچنین این موضوع از صحنه‌های مختلف با چالش رو به رو است.

به طور کلی، مهمترین چالش‌هایی که باعث شده داده کاوی در برنامه‌های مراقبت‌های بهداشتی با سایر برنامه‌ها متفاوت باشد، موارد زیر است:

۳-۲-کیفیت داده‌ها

اولین چالش و شاید مهم‌ترین چالش پیش رو این است که برای دستیابی به اطلاعات مفید و قابل اعتماد از فرآیند داده کاوی، به داده‌هایی با کیفیت خوب نیاز است[10]. کیفیت داده را می‌توان با تعدادی از ابعاد ارزیابی کرد: کامل بودن، اعتبار، انسجام[11]. جاستین و همکاران [12] معیارها و استانداردهای کیفیت داده را از طریق یک جلسه گروهی با مدیران تعیین کرده سپس معیارهای بدست آمده را با استفاده از رگرسیون لجستیک دو جمله‌ای پیاده‌سازی کردند تا همبستگی‌ها و ارتباطات بین معیارها مشخص گردد.

داده‌های بهداشتی معمولاً از پایگاه داده‌های مختلف با طراحی‌های کاملاً متفاوت جمع‌آوری می‌شوند که باعث شده داده‌ها پیچیده، کثیف، دارای تعداد زیادی داده از دست رفته و استانداردهای کدگذاری متفاوت برای زمینه‌های مشابه باشند [13]. اگر یک ارائه دهنده خدمات بهداشتی، به عنوان مثال یک بیمارستان به منظور تجزیه و تحلیل داده‌ها اقدام به جمع‌آوری داده‌ها می‌کند، بایستی یک سری چارچوب‌ها و روش‌ها برای به دست آوردن اطلاعات مفید جهت بهبود کیفیت خدمات و کارآمد ساختن هزینه‌های آنها تعریف کند. این امر به طراحی پایگاه داده بسیار انعطاف پذیر و مقیاس پذیر برای هر واحد نیاز دارد که هر زمان لازم باشد به راحتی بتوان به آن متصل شد و از همان استانداردها برای وارد کردن داده‌ها استفاده کند [10]. آنجلینا و همکاران [14] با استفاده از داده‌های موجود در پایگاه داده MIMIC-III که از کیفیت بالا و همچنین دسترسی آزادنه برخوردار است برای ارزیابی کیفیت داده‌ها بهره بردند. آنها برای ارزیابی از چرخه L* استفاده کردند که شامل ۵ مرحله اصلی بوده: برنامه‌ریزی - استخراج - ایجاد یک مدل کنترل جریان و اتصال آن به گزارش رویداد - ایجاد یک مدل فرآیند یکپارچه - پشتیبانی. آن‌ها بین مرحله اول و دوم یعنی برنامه‌ریزی و استخراج یک مرحله جدید به منظور آماده‌سازی MIMIC-III جهت عملیات فرآیند کاوی، افزودند این مرحله که بازسازی پایگاه داده مذکوراست باعث می‌شود مراحل ۱ تا ۳ به تعداد دفعات مورد نیاز جهت بهبود کیفیت دادها تکرار شود. رابرت و همکاران [15] یک رویکرد فرآیند محور و مبتنی بر کیفیت داده‌ها که برگرفته شده از روش Crisp-DM (Cross Industry Standard Process for Data Mining) است ارائه دادند. آن‌ها با استفاده از روش پیشنهادی مشکلات مربوط به کیفیت داده‌ها را شناسایی کردند. و همچنین ویژگی‌ها و کیفیت داده‌های ورودی در مراحل اولیه فرآیند کاوی را در نظر گرفته و یک ارزیابی کامل روی کیفیت داده‌های پردازش شده انجام دادند. طارق و همکاران [16] برای کاهش محدودیت‌های مربوط به تخیص موضوع و کیفیت داده‌ها یک چارچوب پیشنهاد دادند. چارچوب پیشنهادی شامل ۴ فاز بود: کشف و تشخیص موضوع – جمع‌آوری داده‌ها – آماده‌سازی داده‌ها ارزیابی کیفیت – تجزیه و تحلیل و نتایج. آن‌ها برای بررسی کیفیت داده‌ها از یک ماتریس ارزیابی استفاده کردند. این ماتریس توسط پردازش زبان طبیعی تولید می‌شود. ماتریس مربوطه، پست هر کاربر را در مجموعه داده‌ها بررسی می‌کند تا اطمینان حاصل کند که شامل اصطلاحات مربوط به هستی‌شناسی است. سپس با اختساب نمرات داده شده به پست هر کاربر ارزیابی کیفیت صورت گرفت. در جدول (۱) خلاصه کارهای صورت گرفته در زمینه حل مشکلات کیفیت داده‌ها آورده شده است.

۳-۳-اشتراک و حریم خصوصی داده‌ها:

کیفیت داده‌ها به عوامل مختلفی از جمله حذف داده‌های پر سر و صدا، داده‌های از دست رفته و ... بستگی دارد. برای حفظ کیفیت در داده‌های مراقبت‌های بهداشتی باید تمام اقدامات لازم انجام شود. اشتراک داده‌ها یک چالش مهم دیگر است. نه بیماران و نه سازمانهای بهداشتی علاقه‌ای به اشتراک اطلاعات خصوصی خود ندارند. همچنین حریم خصوصی سوابق و استفاده اخلاقی از اطلاعات بیمار نیز یکی از موانع بزرگ در داده کاوی در مراقبت‌های بهداشتی است. برای اینکه داده کاوی دقیق‌تر باشد، به تعداد قابل توجهی از سوابق واقعی نیاز دارد. سوابق بهداشتی اطلاعات خصوصی هستند و با این حال، استفاده از این پرونده‌های خصوصی ممکن است به جلوگیری از بیماری‌های کشنده کمک کند[6].

برای اینکه بتوانید به داده‌های بهداشتی مورد نیاز برای تجزیه و تحلیل دسترسی پیدا کنید، باید پروتکل‌های امنیتی کافی در سایت انبار داده اجرا شود، تا تحلیلگران بتوانند برای اجرای تجزیه و تحلیل مورد نیاز، به داده‌ها برسند. اگرچه اجرای این پروتکل‌های امنیتی ممکن است آزاردهنده باشد، اما وجود آنها برای هم محافظت از حریم خصوصی بیماران تا آنجا که ممکن است و هم به تحلیلگران امکان دسترسی به داده‌ها برای انجام تجزیه و تحلیل خواهد داد، می‌تواند کمک کننده باشد [10].

یک سری تکنیک وجود دارد که ضمن حفظ اطلاعات کاربران می‌توان از داده‌ها دانش کسب کرد که موسوم به PPDM (Privacy Preserving Data Mining) است و در جدول (۲) آورده شده است. دسته‌بندی های زیادی برای PPDM وجود دارد اما یکی از این دسته‌بندی ها مربوط به کرسو و همکاران [17] است که تکنیک‌ها به ۵ دسته کلی تقسیم‌بندی شده‌اند:

ناشناس‌سازی داده‌ها. نیکونج و همکاران [18] با استفاده از تکنیک ناشناس‌سازی منبع یک طرح جهت حفظ حریم خصوصی در جمع‌آوری و استخراج اطلاعات از داده‌های EHR ارائه دادند. طرح مورد نظر داده‌ها را بدون از دست رفتن هیچ یک از اطلاعات از تمامی سیستم‌های EHR جمع‌آوری کرده و در یک سرور مرکزی ذخیره می‌کند و سپس به تجزیه و تحلیل داده‌ها بدون درگیر شدن EHRها می‌پردازد. در ناشناس‌سازی این نکته حائز اهمیت است که باید توجه داشت میزان ناشناس‌سازی آنقدر زیاد نباشد که دیگر داده‌ها کیفیت اولیه خود را نداشته باشند و مناسب برای تجزیه و تحلیل نباشند. بنابراین ایجاد تعادل بین ناشناس‌سازی امری ضروری و مهم برای محافظت از حریم خصوصی در داده‌ها تحلیلی است [10].

اغتشاش داده‌ها. چامیکارا و همکاران [19] جهت حفظ حریم خصوصی در سیستم‌های فیزیکی سایبری هوشمند چارچوب SEAL را بر

جدول (۱): خلاصه کارهای صورت گرفته در زمینه کیفیت داده ها

محققین (سال انتشار)	مجموعه داده(ها)	الگوریتم(های) استفاده شده	هدف از انجام	منبع
جاستین و همکاران (۲۰۱۷)	شامل ۱۹۶۹۶۷ برخورد فردی با بیمار از پایگاه داده ابزار گزارش مبتنی بر وب	logit binomial regression (LBR)	بررسی کیفیت داده‌های تعاملی بین به موقع بودن، اعتبار، کامل بودن و میزان استفاده میان کاربران مراقبت‌های اولیه	[12]
آنجلینا و همکاران (۲۰۱۹)	۴۶۵۲۰ بیمار از پایگاه داده Medical Information Mart for Intensive Care III (MIMIC-III)	–	ارزیابی کیفیت داده‌های فرآیند کاوی مراقبت‌های بهداشتی با استفاده از MIMIC-III	[14]
رابرت و همکاران (۲۰۱۹)	پایگاه داده های Queensland Ambulance Service (QAS) و Retrieval Services Queensland (RSQ)	–	توسعه یک رویکرد فرآیند محور و مبتنی بر کیفیت داده‌ها برای آماده‌سازی جهت تجزیه و تحلیل فرآیند کاوی با استفاده از روش Crisp-DM	[15]
طارق و همکاران (۲۰۲۰)	–	Natural Language Processing (NLP)	توسعه و ارزیابی چارچوبی برای استخراج و تجزیه و تحلیل محتوای رسانه‌های اجتماعی در ارتباط با سو مصرف مواد مخدر به منظور کاهش محدودیت‌های مربوط به کیفیت داده‌ها و تشخیص موضوع	[16]

اساس درون یابی چیبشف پیشنهاد کردند. نتایج نشان دهنده آن بود که SEAL در حفظ حریم خصوصی داده‌های بزرگ و جریانی طبقه‌بندی شده، عملکرد خوبی دارد.

رمزنگاری. شو و همکاران [20] چارچوبی برای حفظ حریم خصوصی با طراحی سه پروتکل به کمک رمزنگاری همریختی پیشنهاد دادند.که پروتکل اول برای بهینه‌سازی عملکرد استخراج و پروتکل دوم و سوم برای بالا بردن تضمین حریم خصوصی طراحی شدند. سین جی و همکاران [21] مدل DAG را برای حفظ حریم خصوصی با استفاده از محاسبات چند جانبه امن (SMC^{Δ}) پیشنهاد دادند. یانگ و همکاران [22] برای پیش‌بینی خطر ابتلا به بیماری طرح EPDP را پیشنهاد دادند. EPDP شامل سه مرحله: مقدار دهی اولیه – آموزش مدل و پیش‌بینی مدل است، که با رمزنگاری همریختی پیاده‌سازی می‌شود. وانگ و همکاران [23] یک طرح یادگیری مشترک با حفظ حریم خصوصی به نام PCML اجرا کردند که یک مدل تشخیصی برای بهبود نتایج پیش‌بینی بیماری در سرویس‌های تشخیص پزشکی انلاین است. PCML عمدتا از چهار مرحله تشکیل شده است: ۱) مقدار دهی اولیه ۲) رمز گذاری مدل تشخیص محلی ۳) یادگیری مدل مشارکتی ۴) نتیجه یادگیری مشارکتی. همچنین به طور خاص، قبل از ارسال به ابر، همه مدل‌های تشخیص محلی توسط صاحب آنها رمزگذاری شده و بدون رمزگشایی در طی فرایند یادگیری مدل مشارکتی محاسبه می‌شوند. عبدالکرم و همکاران [24] طرح پیشنهادی(PPCDSS) را ارئه دادند. طرح مورد با استفاده از الگوریتم پیشنهادی (PPSDT) و با به کارگیری رمزنگاری همریختی جهت تولید مدل طبقه‌بندی سیستم DSS بدون اشتراک گذاری داده‌ها پیاده‌سازی شده است.

منطق فازی. نیکونج و همکاران [25] روشی برای ادغام سیستم‌های HER محلی بدون اشتراک گذاری داده‌ها ارائه کردند. این روش موسوم به PPDARM، برای حفظ حریم خصوصی اطلاعات بیماران و بالا بردن دقت نتایج قوانین وابستگی در داده‌های تقسیم‌بندی شده افقی به کار گرفته شده است. آنها در پژوهشی دیگر [26] به جای استفاده از داده‌های تقسیم‌بندی شده افقی از داده‌های تقسیم‌بندی شده عمودی برای کشف همبستگی مربوط به بیماری و حفظ حریم خصوصی بیماران استفاده کردند.

شبکه عصبی. مانیکندان و همکاران [27] برای تولید کردن سروصدا جهت حفظ حریم خصوصی داده‌های بهداشتی از شبکه عصبی پیشخور استفاده کردند. آن‌ها شبکه تولید شده که از سه لایه وروری، پنهان، خروجی تشکیل می‌شود را به کمک الگوریتم پس انتشار طراحی کردند. سپس ورودی‌های شبکه با استفاده از الگوریتم پس انتشار به خروجی داده می‌شوند . خروجی حاصل با مقدار هدف مقایسه می‌شود. اگر هدف و خروجی بدست آمده یکسان باشد، خروجی به عنوان سروصدا در نظر گرفته می‌شود. در صورت تناقض، مقدار خروجی دوباره به لایه مخفی انتشار می‌یابد. وزن‌ها در لایه پنهان تنظیم و پردازش می‌شوند. سپس مقدار به لایه خروجی ارسال می‌شود و دوباره همان روند ادامه می‌یابد تا زمانی که هدف و مقدار بدست آمده یکسان شوند. ژانگ و همکاران [28] چارچوب PPDP -privacy) (preserving disease prediction برای ایجاد یک سیستم پیش‌بینی بیماری با حفظ حریم خصوصی داده‌های بهداشتی پیشنهاد کردند. چارچوب پیشنهادی با کمک الگوریتم یادگیری پرسپترون تک لایه پیاده‌سازی می‌شود.

۳-٤-از دست دادن اطلاعات در پیش پردازش:

پیش پردازش داده‌ها، از جمله مدیریت داده‌های از دست رفته، یک از پر هزینه‌ترین و زمانبرترین قسمت‌های فرآیند داده کاوی است. متداول‌ترین روش مورد استفاده در مقاله‌های بررسی شده نادیده گرفتن یا حذف داده‌های از دست رفته بود. در یک مطالعه، تقریباً ۳۳٪ از داده‌ها و ۲۳ پرونده از ۱۳۳

جدول (۲): تکنیک‌های PPDM

تکنیک	محققین (سال انتشار)	نام طرح	مجموعه داده(ها)	هدف از انجام	توضیحات روش پیشنهادی	منبع
ناشناس‌سازی داده‌ها	نیکونج و همکاران (۲۰۲۰)	–	مجموعه داده بیماری‌های قلبی شامل ۳۰۳ پرونده بیمار از پایگاه داده UCI	بهبود خدمات مراقبت‌های بهداشتی با حفظ حریم خصوصی به کمک روش منبع ناشناس	طرح پیشنهادی داده‌ها را از تمام سیستم‌های EHR بدون هیچ گونه از دست دادن اطلاعات جمع‌آوری کرده و در یک سرور داده کاوی مرکزی ذخیره می‌کند. سرور می‌تواند هر تکنیک داده کاوی را که توسط محققان پزشکی یا پزشکان بر روی داده‌های جمع‌آوری شده مراقبت‌های بهداشتی انتخاب شده است، برای کشف الگوهای دقیق مراقبت‌های بهداشتی اجرا کند.	[18]
اغتشاش داده‌ها	چامیکارا و همکاران (۲۰۱۹)	(SEAL)	UCI	طراحی الگوریتم SEAL برای حفظ حریم خصوصی در سامانه فیزیکی–سایبری هوشمند	SEAL یک سیستم اغتشاش خطی که بر اساس درون یابی چند جمله‌ای چیشف ساخته شده است. و شامل سه مرحله جهت پردازش داده‌ها است: تعیین حساسیت مجموعه داده – هدایت درون یابی چند جمله با سروصدای کالیبره شده – تولید داده‌های اغتشاشی به کمک تابع تقریبی.	[19]
رمزنگاری	شو و همکاران (۲۰۱۷)	–	–	حفظ حریم خصوصی در استخراج مجموعه اقلام مکرر بر روی داده‌های ابر رمزگذاری شده	ایجاد یک چارچوب جهت حفظ حریم خصوصی با طراحی سه پروتکل به کمک رمزنگاری همریختی. پروتکل اول برای بهینه‌سازی عملکرد استخراج و پروتکل دوم و سوم برای بالا بردن تضمین حریم خصوصی طراحی شدند.	[20]
	کاور و همکاران (۲۰۱۸)	(PPCF)	۱- داده‌های مراقبت‌های بهداشتی ۱۰۰۰۰ بیمار از ۵۰۰ پزشک ۲- مجموعه داده MovieLens شامل ۹۴۳ کاربر برای ۱۶۸۲ فیلم	حفظ حریم خصوصی در سیستم‌های توصیه گر مراقبت‌های بهداشتی به کمک پالایش گروهی	در مدل مورد نظر از سه پروتکل استفاده شده است: ۱- شباهت بین موارد توسط پروتکل اول ایجاد می‌شود ۲- محاسبه طول بردار موارد را پروتکل دوم انجام می‌دهد ۳- پیش‌بینی ها توسط پروتکل سوم تولید می‌شود.	[32]
	سین جی و همکاران (۲۰۱۸)	(DAG)	دو مجموعه داده از UCI اولی شامل ۱۵۹۸ تاپل و دومی ۲، ۰۴۹، ۲۸۰ تاپل	ایجاد چارچوبی برپایه SMC برای حفظ حریم خصوصی	مدل مورد نظر شامل ۷ عملگر پایه است و دارای سه نوع گره است: منبع که ورودی مدل، سینک که خروجی مدل و عملگر هم ۷ عملیات پایه است. همچنین مدل با کمک پروتکل SMC طراحی شده است	[21]
	یانگ و همکاران (۲۰۱۹)	(EPDP)	مجموعه داده AID شامل ۱۲۰ نمونه از پایگاه UCI و مجموعه داده ترکیبی تولید شده از ۵۰۰ تاپل	ارائه طرح EPDP جهت پیش‌بینی بیماری در مراقبت‌های بهداشتی با حفظ حریم خصوصی	سیستم دارای دو مرحله است یکی آموزش مدل و دیگری پیش‌بینی مدل که هر دو از راه دور انجام می‌شود. در مرحله آموزش مدل از طبقه‌بندی کننده بیز ساده استفاده شده است. سپس در مرحله پیش‌بینی با استفاده از داده ساختار فیلتر بلوم به تشخیص بیماری بیماران کمک می‌کند.	[22]
	وانگ و همکاران (۲۰۱۹)	(PCML)	مجموعه داده بیماری تیروئید با ۳۱۷۵ مورد و مجموعه داده بیماری‌های قلب با ۸۹۸ مورد	بهبود نتایج پیش‌بینی بیماری در سرویس‌های تشخیص پزشکی انلاین با حفظ حریم خصوصی	در ابتدا پارامترهای سیستم توسط مرجع قابل اعتماد تولید شده سپس با استفاده از سیستم رمزنگاری paillier کلید عمومی PK و خصوصی SK محاسبه شده و کلید SK را به چند قسمت برای مراکز بهداشتی و سرور ابری تقسیم می‌کند. سپس مراکز بهداشتی مدل‌های تشخیصی خود را رمزگذاری کرده و به سرور ارسال می‌کنند و سرور پس از محاسبات، نتایج بدست امده را به مراکز بهداشتی ارسال کرده و	[23]

					آنها بعد از رمزگشایی می‌توانند مدل تشخیصی را بدست آورند	
	عبدالکرم و همکارن (۲۰۱۹)	(PPCDSS)	مجموعه داده التهاب حاد از پایگاه داده UCI	طبقه‌بندی علائم بیماران در سیستم پشتیبانی تصمیم با حفظ حریم خصوصی	مدل مورد نظر بر اساس الگوریتم پیشنهادی PPSDT است که از ترکیب الگوریتم C4.5 با رمزنگاری همریختی paillier به وجود می‌آید	[24]
منطق فازی	نیکونج و همکاران (۲۰۱۸)	(PPDARM)	مجموعه داده‌های بیماری‌های قلبی UCI شامل ۳۰۳ بیمار	حل مشکل استخراج قوانین وابستگی بدون اشتراک گذاری داده‌های محلی EHR	رویکرد مورد نظر حداقل به سه شرکت کننده نیاز دارد. فرض بر این است که کانال‌های ارتباطی بین شرکت کنندگان ناامن هستند. هر دشمن می‌تواند تمام ارتباطات بین شرکت کنندگان را ذخیره کند تاp-value برخی دیگر از شرکت کنندگان را بیاموزد. برای جلوگیری از این مشکل از الگوریتم محساباتی دیفی هلمن (CDH^F) استفاده شده است. روش پیشنهادی براساس این فرض که حل مسئله CDH از نظر محاسباتی سخت است، ایمن است.	[25]
	نیکونج و همکاران (۲۰۱۹)	–	پایگاه داده معاینه سلامت یک بیمارستان محلی در تایوان شامل ۱۲۲۶۵۸ مورد	حفظ حریم خصوصی در هنگام اشتراک گذاری داده‌های بهداشتی جهت یکپارچه‌سازی سیستم EHR	-	[26]
شبکه عصبی	مانیکاندان و همکاران (۲۰۱۷)	–	مجموعه داده EHR	حفظ حریم خصوصی داده‌های بهداشتی با استفاده از شبکه عصبی	ورودی‌های شبکه با استفاده از الگوریتم پس انتشار به خروجی داده می‌شوند . خروجی حاصل با مقدار هدف مقایسه می‌شود. اگر هدف و خروجی بدست آمده یکسان باشد، خروجی به عنوان سروصدا در نظر گرفته می‌شود. در صورت تناقض، مقدار خروجی دوباره به لایه مخفی انتشار می‌یابد. وزن‌ها در لایه پنهان تنظیم و پردازش می‌شوند. سپس مقدار به لایه خروجی ارسال می‌شود و دوباره همان روند ادامه می‌یابد تا زمانی که هدف و مقدار بدست آمده یکسان شوند.	[27]
	ژانگ و همکاران (۲۰۱۸)	(PPDP)	مجموعه داده‌های سرطان سینه شامل ۶۸۳ نمونه و بیماری‌های قلبی شامل ۲۹۷ نمونه از UCI	توسعه یک سیستم پیش‌بینی بیماری با حفظ اطلاعات بیماران به کمک چارچوب PPDP مبتنی بر سرور ابری	چارچوب پیشنهادی بدین شکل عمل می‌کند که بیمار برای پیش‌بینی بیماری خود علائم را به بیمارستان ارسال کرده سپس بیمارستان علائم رمزگذاری شده را به سرور ابری فرستاده و ابر از مدل‌های پیش‌بینی آموزش دیده رمزگذاری شده برای تشخیص استفاده خواهد کرد و به بیمارستان ارسال می‌کند. در نهایت بیمارستان نتایج پیش‌بینی را به بیمار بر می‌گرداند	[28]

پرونده به دلیل ثبت نشدن در پایگاه داده حذف شدند [12]. با حذف موارد با ارزش مفقود شده و هزینه‌های اضافی، مقدار قابل توجهی از اطلاعات از دست می‌رود. تحقیق‌های آینده باید یافتن روشی بهتر برای برآورد ارزش از دست رفته نسبت به حذف باشد. علاوه بر این، تکنیک‌های جمع‌آوری داده باید توسعه داده یا اصلاح شوند تا از این مسئله جلوگیری شود. مشابه داده‌های از دست رفته، حذف یا نادیده گرفتن روشی متداول برای مدیریت داده‌های پرت است. تحقیق‌های آینده باید به جای چشم پوشی از داده‌های پرت، آنها را تجزیه و تحلیل کرده تا بینش پیدا کند[13].

۳-۵-تکیه بر مدلهای پیش‌بینی کننده

نباید انتظارات غیرواقعی از مدل‌های داده کاوی ساخته شده وجود داشته باشد. هر مدل دقتی دارد. وقتی یک مدل پیش‌بینی برای تشخیص یک مسئله

سلامتی به عنوان مثال تشخیص نوع غده تیروئید ساخته شده است، معمولاً ۱۰۰٪ دقت ندارد. بسته به میزان دقت مدل و اهمیت تصمیم‌گیری با استفاده از آن مدل، باید تصمیم بگیریم که چقدر می‌توانیم به نتایج این مدل داده کاوی اعتماد کنیم. به خصوص هنگامی که ما با تصمیمات پزشکی بالینی مبتنی بر نتایج یک یا چند مطالعه داده کاوی سر و کار داریم، مهم است که در نظر بگیریم که تنها اتکا به مدل‌های پیش‌بینی هنگام تصمیم گیری‌های حیاتی که مستقیماً بر زندگی بیمار تأثیر می‌گذارد، خطرناک خواهد بود و این را نباید از مدل پیش‌بینی انتظار داشت.

با این حال، مدل‌های داده کاوی و به طور خاص مدل‌های پیش‌بینی هنگامی که به عنوان نظر دوم برای تصمیم پزشک در مورد درمان استفاده می‌شوند، می‌توانند بسیار مفید باشند. در برخی موارد، پزشک ممکن است شدت بیماری بیمار را کم ارزیابی کند و مدل ساخته شده به وی هشدار می‌دهد تا برخی از جنبه‌هایی را که ممکن است از دست رفته باشد و بیمار را در معرض خطر قرار دهد، در نظر بگیرد. از طرف دیگر، اگر پزشک از اینکه بیمار نیاز به جراحب دارد یا خیر این مدل‌ها می‌توانند در تشخیص صحیح به پزشک کمک کنند [10].

۳-۶-انواع روش‌ها و ریاضیات پیچیده

از آنجا که ریاضیات پایه‌ای تقریباً همه تکنیک‌های داده کاوی پیچیده است و برای افراد غیر فنی به راحتی قابل درک نیست، بنابراین، پزشکان و اپیدمیولوژیست‌ها معمولاً ترجیح می‌دهند که با روش‌های سنتی آمار ادامه کار دهند. علاوه بر این، از آنجا که تکنیک‌ها و روش‌های متفاوت در داده کاوی وجود دارد، برای پزشکان دشوار است که با تمام روش‌های مختلف آشنا شوند و به راحتی روش مناسب را انتخاب کنند.

در این مورد، لازم است از تحلیلگرانی که با تکنیک‌های مختلف داده کاوی آشنایی بیشتری دارند کمک گرفته شود تا بتوانند روش‌های مختلفی را به کار گیرند و از قدرت داده کاوی در برخورد با مجموعه‌های داده بزرگ و پذیرش ویژگی‌های ورودی بیشتر در مقایسه با آمار سنتی استفاده کنند[10].

۳-۷-ادغام در سیستم مراقبت‌های بهداشتی

تعداد کمی از تحقیق‌های انجام شده تلاش کردند تا فرآیند داده کاوی را در چارچوب تصمیم‌گیری واقعی ادغام کنند. تأثیر کشف دانش از طریق داده کاوی بر حجم کار و زمان حرفه‌ای مراقبت‌های بهداشتی مشخص نیست. مطالعه‌های آینده باید یکپارچه‌سازی سیستم توسعه یافته را در نظر بگیرند و تأثیر آن را بر محیط کار بررسی کنند[13].

۳-۸-اتوماسیون فرآیند داده کاوی برای کاربران غیر متخصص

کاربران نهایی داده کاوی در مراقبت‌های بهداشتی پزشکان، پرستاران و متخصصان مراقبت‌های بهداشتی هستند که آموزش محدودی در زمینه تجزیه و تحلیل دارند. یک راه حل برای این مشکل توسعه یک سیستم خودکار (یعنی بدون نظارت انسانی) برای کاربران نهایی است. تانیا و همکاران [29] معماری ADA-HEALTH را معرفی کردند که بر پایه تجزیه و تحلیل اطلاعات داده‌های بهداشتی با حداقل دخالت کاربر است. آنها با استفاده از روش خوشه‌بندی K میانگین به متوسط دقت ۷۹.۷۲٪ رسیدند.

راه دیگر این است که می‌توان یک ساختار خودکار مبتنی بر ابر برای جلوگیری از خطاهای پزشکی ایجاد کرد [28]. اما این وظیفه چالش برانگیز است زیرا شامل زمینه‌های مختلف کاربردی است و یک الگوریتم برای همه برنامه‌ها دقت مشابهی نخواهد داشت[13].

۳-۹-ماهیت میان رشته‌ای مطالعه و دانش تخصصی حوزه

تجزیه و تحلیل بهداشت و درمان یک زمینه تحقیقاتی بین رشته‌ای است. کشف دانش از داده‌های پزشکی یک کار دشوار است زیرا داده‌های موجود پر سر و صدا، بی ربط و گسترده نیز هستند[30]. به عنوان نوعی تجزیه و تحلیل، داده کاوی باید در ترکیب با نظر متخصص از حوزه‌های خاص مراقبت‌های بهداشتی و مشکل خاص (به عنوان مثال انکولوژیست برای مطالعه سرطان، متخصص قلب برای CVD) استفاده شود. یکی از پژوهش‌های صورت گرفته در این راستا مربوط به آنتونی و همکاران [31] است که به طور خاص برای ایجاد یک شبکه بیز از متغیرهای متخصص دامنه استفاده کرده‌اند و همچنین روشی را برای استخراج قضاوت متخصص ارائه کردند. تقریباً ۳۲٪ از مقاله‌های تحلیلی از نظر متخصص به هیچ وجه استفاده نکردند. تحقیق‌های آینده باید شامل اعضای رشته‌های مختلف از جمله مراقبت‌های بهداشتی باشد[13].

۳-۱۰-عدم توانایی تغییر عادات پزشکان

حتی اگر نتایج داده کاوی معتبر باشد، متقاعد کردن پزشکان برای تغییر عادت‌های خود بر اساس شواهد ممکن است مشکل بزرگتری باشد. چند مورد مشاهده شده که پزشکان بیمارستان حتی در مواجهه با شواهد از تغییر سیاست بیمارستان خودداری می‌کنند. در یک مورد مشخص شد که پزشکان از کالبد شکافی بدون شستن دست بیرون می‌آیند و منجر به احتمال بالای مرگ در بیمارانی شده که پس از کالبد شکافی آنها را معالجه می‌کنند. با ارائه این شواهد، پزشکان هنوز از تغییر عادت‌های خود سر باز زدند[13].

۳-۱۱-تفاوت داده کاوی سنتی و پزشکی

در تحقیق‌های پزشکی، داده کاوی با یک فرضیه آغاز می‌شود و سپس نتایج متناسب با آن فرضیه تنظیم می‌شود. این فرآیند از عمل استاندارد داده کاوی، که به سادگی با مجموعه داده‌ها و بدون فرضیه شروع می‌شود، متفاوت است.

همچنین، داده کاوی سنتی نگران الگوها و روند در مجموعه داده‌ها است، در حالی که داده کاوی در پزشکی بیشتر به اقلیتی علاقه‌مند است که با الگوها و روندها مطابقت نداشته باشد. علاوه بر این، بیشتر نتایج استاندارد داده کاوی مربوط به توصیف اما عدم توضیح الگوها و روندها است. در مقابل، پزشکی به این توضیحات نیاز دارد زیرا یک تفاوت جزئی می‌تواند تعادل بین مرگ و زندگی را تغییر دهد[6].

٤- نتیجه گیری

با توجه به رشد روز افزون بهره مندی از دانش داده کاوی در حوزه سلامت که منجر به نتایج بهتر شده، مشکلاتی را با خود همراه داشته که منجر به بدست نیامدن نتایج مطلوب گردیده است. تکنیک‌های به کار گرفته شده در مطالعه‌های اخیر حاکی از آن بوده که دقت مدل‌های پیش‌بینی متناسب با هر تکنیک متفاوت است. با توجه به این موضوع دقت هیچ کدام از مدل‌ها ۱۰۰ درصد نیست و از آن جایی که حوزه سلامت، حوزه‌ای بسیار حساس و کوچک‌ترین اشتباه در پیش‌بینی انجام شده به جان انسان‌ها ختم شده، مدل‌های پیش‌بینی کننده در اولویت‌های بعدی برای استفاده قرار می‌گیرند. دلایل متعددی برای عدم کارکرد صحیح این مدل‌ها وجود دارد که دربخش ۳ به آن اشاره گردید.

همچنین با توجه به یکی از بزرگ‌ترین چالش‌های پیش رو که کیفیت داده‌های جمع‌آوری شده است بایستی زمینه همکاری بین سازمان‌های بهداشتی و تحلیل گران بیشتر شود تا بتوان داده‌هایی با کیفیت برای عمل تجزیه و تحلیل در اختیار مهندسین داده قرار داد. بایستی پایگاه داده‌های منسجم و یکپارچه که امکان به اشتراک گذاری اطلاعات وجود داشته باشد، ایجاد کرد. همچنین سازمان‌های بهداشتی همکاری بیشتری با سازمان‌های جمع‌آوری کننده اطلاعات بهداشتی داشته باشند تا داده‌های با کیفیت در اختیار مهندسین تحلیل گر قرار دهند. موردی که کمتر به آن در تحقیق‌های صورت گرفته توجه شده است مسئله استفاده از دانش متخصص حوزه می‌باشد که باعث شده نتایج بدست آمده مورد رضایت نبوده و به کار گرفته نشوند.

داده‌های پزشکی بر اثر دو عامل مهم از کیفیت پایینی برخودار هستند: اول عدم همکاری سازمان‌های بهداشتی با پایگاه داده‌های جمع‌آوری کننده اطلاعات که به آن اشاره گردید و مورد دوم عدم حفظ حریم خصوصی می‌باشد که در این زمینه فعالیت‌هایی صورت گرفته است اما همچنان به طور کامل مشکل حریم خصوصی را برطرف نکرده و این عامل باعث شده تا سازمان‌های بهداشتی میل و رغبت کمتری نسبت به دادن اطلاعات بیماران به تحلیلگران داده نشان دهند.

داده کاوی در پزشکی هنوز در مراحل اولیه خود به سر می‌برد و به بلوغ لازم نرسیده است و راه برای پیشرفت آن زیاد است. تحقیق‌های آینده بایستی به مسائل مطرح شده توجه ویژه‌ای داشته باشند تا بتوان دقت مدل‌های پیش‌بینی کننده را بالا برد و دربخش‌های مختلف به بهترین شکل ممکن از این مدل‌ها استفاده کرد.

مراجع

[1] E. M. Beulah *et al.*, "Application of Data mining in healthcare: A survey," *Asian Jr. of Microbiol. Biotech. Env. Sc.*, vol. 18, no. 4, 2016.

[2] I. Ţăranu, "Data mining in healthcare: decision making and precision," *Database Systems Journal*, vol. 6, no ,۴ . pp. 33-40, 2016.

[3] M. N. Almunawar and M. Anshari, "Health information systems (HIS): Concept and technology," *arXiv preprint arXiv:1203.3923*, 2012.

[4] W. H. Organization, *Framework and standards for country health information systems*. World Health Organization, 2008.

[5] M. Mohd, A. Papazafeiropoulou, R. J. Paul, and L. K. Stergioulas, "Investigating evaluation frameworks for health information systems," *Int. J. Medical Informatics*, vol. 77, no. 6, pp. 377-385, 2008.

[6] N. Soni and C. Gandhi, "Application of data mining to health care," *International Journal of Computer Science and its Applications*, 2010.

[7] N. Jothi and W. Husain, "Data mining in healthcare–a review," *Procedia computer science*, vol. 72, pp. 306-313, 2015.

[8] P. Ahmad, S. Qamar, and S. Q. A. Rizvi, "Techniques of data mining in healthcare: a review," *International Journal of Computer Applications*, vol. 120, no. 15, 2015.

[9] I. Yoo *et al.*, "Data mining in healthcare and biomedicine: a survey of the literature," *Journal of medical systems*, vol. 36, no. 4, pp. 2431-2448, 2012.

[10] M. H. Tekieh and B. Raahemi, "Importance of data mining in healthcare: a survey," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* ,۲۰۱۵ ,pp. 1057-1062 .

[11] N. Garcelon, A. Burgun, R. Salomon, and A. Neuraz, "Electronic health records for the diagnosis of rare diseases," *Kidney International*, vol. 97, no. 4, pp. 676-686, 2020.

[12] J. St-Maurice and C. Burns, "An exploratory case study to understand primary care users and their data quality tradeoffs," *Journal of Data and Information Quality (JDIQ)*, vol. 8, no. 3-4, pp. 1-24, 2017.

[13] M. S. Islam, M. M. Hasan, X. Wang, and H. D. Germack, "A systematic review on healthcare analytics: application and theoretical perspective of data mining," in

[26] N. Domadiya and U. P. Rao, "Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data," *Procedia computer science*, vol. 148, pp. 303-312, 2019.

[27] G. Manikandan, N. Sairam, M. Sathya Priya, and H. V. Sri Radha Madhuri, "A new approach for ensuring medical data privacy using neural networks.۲۰۱۷ ",

[28] C. Zhang, L. Zhu, C. Xu, and R. Lu, "PPDP: An efficient and privacy-preserving disease prediction scheme in cloud-based e-Healthcare system," *Future Generation Computer Systems*, vol. 79, pp. 16-25, 2018.

[29] T. Cerquitelli, E. Baralis, L. Morra, and S. Chiusano, "Data mining for better healthcare: A path towards automated data analysis?," in *2016 IEEE 32nd International Conference on Data Engineering Workshops (ICDEW)*, 2016: IEEE, pp. 60-63 .

[30] M. Durairaj and V. Ranjani, "Data mining applications in healthcare sector: a study," *International journal of scientific & technology research*, vol. 2, no. 10, pp. 29-35, 2013.

[31] A. C. Constantinou, N. Fenton, and M. Neil, "Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved," *Expert systems with applications*, vol. 56, pp. 197-208, 2016.

[32] H. Kaur, N. Kumar, and S. Batra, "An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system," *Future Generation Computer Systems*, vol. 86, pp. 297-307, 2018.

زیرنویس‌ها

[1] Health Information System
[2] Electronic Health Record
[3] Information and Communications Technology
[4] Knowledge Discovery in Databases
[5] Secure multi-party computation
[6] Computational Diffie–Hellman

Healthcare, 2018, vol. 6, no. 2: Multidisciplinary Digital Publishing Institute, p. 54 .

[14] A. P. Kurniati, E. Rojas, D. Hogg, G. Hall, and O. A. Johnson, "The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database," *Health informatics journal*, vol. 25, no. 4, pp. 1878-1893, 2019.

[15] R. Andrews *et al.*, "Leveraging data quality to better prepare for process mining: an approach illustrated through analysing road trauma pre-hospital retrieval and transport processes in Queensland," *International journal of environmental research and public health*, vol. 16, no. 7, p. 1138, 2019.

[16] T. Nasralah, O. El-Gayar, and Y. Wang, "Social media text mining framework for drug abuse: Development and validation study with an opioid crisis case analysis," *Journal of medical Internet research*, vol. 22, no. 8, p. e18350, 2020.

[17] I. Kreso, A. Kapo, and L. Turulja, "Data mining privacy preserving: Research agenda," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1392, 2020.

[18] N. Domadiya and U. P. Rao, "Improving healthcare services using source anonymous scheme with privacy preserving distributed healthcare data collection and mining," *Computing*, pp. 1-23, 2020.

[19] M. A. P. Chamikara, P. Bertók, D. Liu, S. Camtepe, and I. Khalil, "An efficient and scalable privacy preserving algorithm for big data and data streams ",*Computers & Security*, vol. 87, p. 101570, 2019.

[20] S. Qiu, B. Wang, M. Li, J. Liu, and Y. Shi, "Toward practical privacy-preserving frequent itemset mining on encrypted cloud data," *IEEE Transactions on Cloud Computing*, 2017.

[21] S. G. Teo, J. Cao, and V. C. Lee, "Dag: A general model for privacy-preserving data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 1, pp. 40-53, 2018.

[22] X. Yang, R. Lu, J. Shao, X. Tang, and H. Yang, "An efficient and privacy-preserving disease risk prediction scheme for e-healthcare," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3284-3297, 2018.

[23] F. Wang *et al.*, "Privacy-preserving collaborative model learning scheme for E-healthcare," *IEEE Access*, vol. 7, pp. 166054-166065, 2019.

[24] A. Alabdulkarim, M. Al-Rodhaan, T. Ma, and Y. Tian, "PPSDT: A novel privacy-preserving single decision tree algorithm for clinical decision-support systems using IoT devices," *Sensors*, vol. 19, no. 1, p. 142, 2019.

[25] N. Domadiya and U. P. Rao, "Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases," *Sādhanā*, vol. 43, no. 8, pp. 1-9, 2018.