یازدهمین کنگره ملی سراسری
فناوریهای نوین در حوزه توسعه پایدار ایران
11th National Congress of
the New Technologies in Sustainable Development of Iran
senaconf.ir

کاربردهای یادگیری ماشین و کلان داده در حوزه مراقبت های بهداشتی

پرویز قربانزاده[۱]، مهدی زینالی[۲]، زهرا تقی پور[۳]

[۱]مدرس دانشگاه, گروه مهندسی فناوری اطلاعات, دانشکده فناوری های صنعتی, دانشگاه صنعتی, ارومیه, ایران؛ P.ghorbanzadeh@uut.ac.ir

[۲] دانشجوی کارشناسی مهندسی کامپیوتر گرایش نرم افزار، دانشگاه آزاد اسلامی, ارومیه, ایران؛ Mahdizeynali٤٢٢٨@gmail.com

[۳]دانشجوی کارشناسی ارشد مهندسی کامپیوترگرایش نرم افزار، دانشگاه آزاد اسلامی واحد ارومیه، ایران, Tagepour.zahra@gmail.com

چکیده

کلان داده درواقع اصطلاحی است که برای حجم زیادی از داده‌ها استفاده می‌شود که خیلی سریع و پیچیده هستند و پردازش آنها با استفاده از روش‌های سنتی، سخت و غیرممکن است. مراقبت های بهداشتی زمینه مناسبی برای کاربردهای یادگیری ماشین و کلان داده در داده های مراقبت های بهداشتی است. پیاده سازی مراقبت های بهداشتی با کلان داده و یادگیری ماشینی الزامات سلامت مشتری را افزایش می دهد. در حال حاضر با توجه به اینکه برنامه های الکترونیکی پرونده سلامت در حال افزایش است، ضروری است که بر استفاده از داده های تولید شده توسط این برنامه ها تمرکز شود. حجم زیادی از داده ها در مراقبت های بهداشتی وجود دارد که مربوط به حوزه های مختلف سلامت به ویژه در حوزه بیمارهای عصبی و قلبی است. این داده ها نیاز به تمرکز ویژه ای داشته و معماری هایی که در حال حاضر بر روی این بخش ها متمرکز می شوند بایستی از آخرین فناوری های روز برای پیش بینی برخی الگوها استفاده نمایند. در این پژوهش، بر پیاده سازی معماری های مختلف مراقبت های بهداشتی تمرکز شده است که از داده های لایو جمع آوری شده از منابع مختلف در سراسر جهان استفاده می نماید. ضمن اینکه، از ترکیب رویکردهای یادگیری ماشین و چارچوب کلان داده برای طراحی یک مدل پیش بینی و تکنیک های دستکاری داده ها استفاده شده است.

واژه‌های کلیدی

یادگیری ماشین، کلان داده، مراقبت های بهداشتی، مدل پیش بینی

۱. مقدمه

تجزیه و تحلیل کلان داده نقش مهمی در مدیریت مدل های پیش بینی ایفا می نماید و داده های به دست آمده از منابع مختلف باعث ایجاد آشفتگی های امنیتی و شکست معماری در مدیریت داده های لایو می شود. در این پژوهش سعی شده است انواع مختلف مدل های داده ای برای پیاده سازی مدل های پیش بینی در حوزه پزشکی توضیح داده شود. تحقیقاتِ پزشکی در حال حاضر بیشتر بر این تمرکز دارد که کدام راه‌حل‌های درمانی بر روی کدام بیماری‌ها و کدام گروه از بیماران بهترین جواب را می‌دهد. به این صورت که یک آزمایش بر روی گروهی از بیماران انجام می‌شود و علائم و نتایجی حاصل می‌شود. اگر این نتایج امیدبخش بود، این آزمایشات برای گروه بزرگ‌تری از بیماران انجام می‌شود. با اینکه پزشکان بایستی آزمایشات را به دقت انجام دهند ولی همواره ریسک خطا وجود دارد. ممکن از راه‌حل درمانی برای گروهِ کوچکی از بیماران خوب جواب بدهد ولی وقتی این راه‌حل برای گروهِ بزرگ‌تری از بیماران انجام می‌شود، عواقبِ بدی برای بعضی از بیماران داشته باشد. در واقع خاصیت عمومی‌سازی ممکن است به خوبی جواب ندهد و این یکی از دلایلِ کندیِ عملیاتِ آزمایش و ریسک بالای آن می‌شود. در این‌جاست که تکنیک‌های کلان‌داده‌ای می‌توانند کاربردی باشند. در واقع با کاویدن داده‌های پزشکی مبتنی بر تمرین برای مثال درمان‌هایی که بر روی هر بیمار انجام شده‌است و مشخصات و ویژگی‌های آن فردِ بیمار می‌توان به اطلاعات مفیدتری دست یافت.

نحوه اعتماد و پردازش داده های جمع آوری شده، رویکرد اصلی مورد نیاز برای تمرکز بر اجرای هر نوع مدل پیش بینی است. به عنوان یک دانشمند داده، جمع آوری و مدیریت داده ها از منابع مختلف و پیش پردازش اطلاعات و استفاده مجدد از اطلاعات یکسان برای اهداف مختلف کاری چالش برانگیز می باشد. آنچه این پژوهش بر آن تمرکز نموده است در طی دو مرحله می باشد. مرحله اول به بررسی رویکردهای مختلفی که در حال حاضر در حوزه پزشکی با کلان داده انجام می شود، می پردازد. در رویکرد دوم، به توضیح معماری های مختلف در چارچوب داده های بزرگ برای رسیدگی به اطلاعات پزشکی پرداخته می شود. همچنین، بر معماری های پیشنهادی تمرکز می شود که نتایج بهتری را در مدیریت داده ها از منابع مختلف جهانی نشان می دهد. تجزیه و تحلیل کلان داده ها مشکل بزرگی است، جهان با آن مواجه است و داده های روزمره جمع آوری شده در طی یک روز تا چندین پتابایت می رسد که ۷۵ درصد از داده ها متعلق به حوزه پزشکی است. مشکلات و سناریوهای مختلفی در حوزه پزشکی عمدتا با کلان داده قابل درک می باشد. کلان داده همراه با سیستم چهار V، در حوزه پزشکی به شرح زیر توضیح داده شده است [۱]

۲. حجم زیاد کلان داده پزشکی

حجم عظیمی از داده ها در حوزه پزشکی ایجاد می شود که از همه انواع دیگر داده های ایجاد شده در هر حوزه ای پیشی می گیرد. تحقیقات در صنعت مراقبت های بهداشتی و تحقیقات DNA در هر ثانیه داده های بیشتری را در قالب متن و تصویر ایجاد می کند [۲]. در این سناریو نحوه ایجاد این نوع کلان داده در زمان کوتاه و همچنین نوع داده ها مورد توجه قرار می گیرد. در مورد بیماری خاصی مانند COVID -۱۹ هیچ اطلاعات پیش زمینه ای ندارد و برای گرفتن نمونه های DNA از موارد مثبت، لازم است تحقیقات بیشتری با ترکیبات مختلف انجام داد که اینکار سبب می شود داده بیشتری ایجاد شده که می تواند به صورت الگو مورد استفاده قرار گیرد. به همین ترتیب، الگوهای مختلف در DNA و تحقیقات متفاوتی که در زندگی روزمره ما انجام می شود، سبب تولید داده های بیشتری می شوند[۳].

۳. سرعت زیاد کلان داده پزشکی

درک چگونگی سرعت ایجاد داده ها برای مدیریت چارچوب و آماده ساختن آن برای پذیرش این نوع جریان کلان داده مهم است. این نوع جریان کلان داده می تواند باعث ازدحام ترافیکی شود. در این میان، پیاده سازی اصلی در چارچوبی است که برای مدیریت داده ها استفاده می شود. ابزارهای مختلف در کلان داده ها می توانند این نوع مسائل مرتبط را برطرف نمایند.

۱.۳. تنوع زیاد کلان داده پزشکی

کلان داده پزشکی از تنوع زیادی برخوردار است. در این سناریو، اطلاعات جمع آوری شده از منابع مورد اعتماد و نوع داده ها نیز مهم است. سه نوع داده در کلان داده پزشکی موجود است :

۱.۴ داده های دارای ساختار پزشکی

این قالب سیستماتیک بوده و می تواند در قالب یک ساختار از پیش طراحی شده با سطرها و ستون های داده باشد. داده‌هایی که می‌توان برای پیش‌بینی استفاده کرد، بیشتر به صورت متن، و در قالب معنایی خواهد بود و می‌توان از آن برای پیش‌بینی موردی در رابطه با مساله استفاده کرد.

۲.۴. داده های بدون ساختار پزشکی

داده های فاقد ساختار در قالب چند رسانه ای بوده و می توانند در قالب های مختلف مانند فیلم ، صدا و تصویر باشند. این داده ها از منابع مختلف در تحقیقات فعلی مانند اسکن MRI برای بیماری های مغز، بدست آمده است. سناریوهای مختلفی برای جمع آوری داده های فاقد ساختار مانند چند رسانه ای برای تجزیه و تحلیل وجود دارد.

۳.۴. داده های نیمه ساختار یافته در پزشکی

اطلاعات نیمه ساختار یافته ترکیبی از هر دو قالب داده های ساختار یافته و فاقد ساختار است. در این سناریو، XML یا JSON به عنوان فرمت های اصلی جمع آوری داده ها استفاده می شود. در این رویکرد، چگونگی جمع‌آوری انواع مختلف اطلاعات به طور همزمان بدون آسیب رساندن به معماری الگوی داده‌ها متمرکز شده است.

۵. صحت داده های پزشکی

قابلیت اعتماد داده های جمع آوری شده از منابع مختلف در این بخش مورد بحث قرار گرفته است. با توجه به اینکه نمی شود روش های آزمایش و خطا را برای نوع داده های لایو انجام داد، در صورت وجود ناسازگاری یا عدم دسترسی به اطلاعات، مدل های تصمیم گیری آسیب دیده و درست عمل نخواهند کرد. یک بررسی استاندارد در مورد استفاده از کلان داده در مراقبت های بهداشتی با استفاده از سه مدل پیش بینی قوی مبتنی بر تحلیل توصیفی، تحلیل پیش بینی کننده و تحلیل تجویزی انجام گرفته است که که رویکرد حاصل از بررسی در شکل ۱ نشان داده شده است.[۱].

شکل ۱: تجزیه و تحلیل مبتنی بر رشد کلان داده در بخش مراقبت های بهداشتی

تحلیل‌های پیش‌بینی کننده (Predictive Analytics) یک روش تحلیلی مهم است که توسط بسیاری از سازمان‌ها برای تعیین ریسک، شناسایی روند کسب و کار در آینده و تشخیص نیازهای پیش روی سازمان استفاده می‌شود. دانشمندان داده با استفاده از داده‌های تاریخی به عنوان منبع خود، از تحلیل رگرسیون و تکنیک‌های یادگیری ماشین برای شناسایی الگوها و روند موجود در آن داده‌ها استفاده می‌کنند. هدف اصلی تحلیل‌های پیش‌بینی کننده، شناسایی، با درجه بالایی از احتمالات و آنچه در آینده اتفاق خواهد افتاد است. این موارد را از تجزیه و تحلیل توصیفی (Descriptive Analytics)، که به تحلیلگران کمک می‌کند تا آنچه قبلاً اتفاق افتاده است، در کنار تجزیه و تحلیل تجویزی (Prescriptive Analytics) را که با استفاده از نرم افزار بهینه‌سازی، بهترین تصمیمات را برای مقابله با روند آشکار شده توسط تجزیه و تحلیل‌های پیش‌بینی کننده، به دست آورده و از هم جدا می‌کنند.

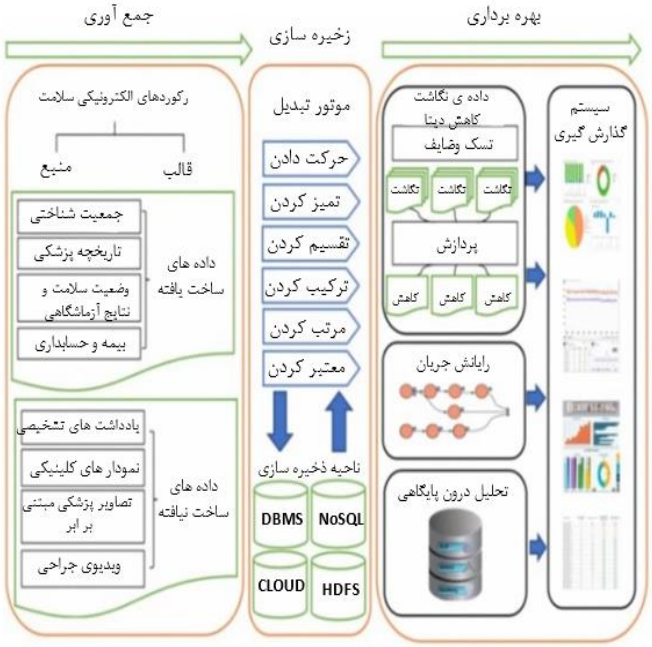
۱.۵. مروری بر ادبیات تحقیق

یکی از مهمترین کاربردهای کلان داده در حوزه سلامت و پزشکی است. به کمک داده‌کاوی و تحلیل دقیق داده‌ها می‌توان به اطلاعات تعداد زیادی از بیماران در سنین مختلف و گروه‌های متنوع رسید و برای مقابله با بیماری‌ها و پیشگیری از ابتلای به آنها، راهکارهای مناسب و کاربردی ارائه داد.با استفاده از کلان داده در زمینه بهداشت و درمان، می‌توان کاری کرد که خدمات بهتری را به مردم ارائه گردد. کاربرد کلان داده در پزشکی شناسایی روش های شخصی سازی شده برای بیماران است. همین شخصی سازی است که باعث افزایش سلامت در جامعه شده و همچنین هزینه های دولت ها را نیز در بخش بهداشت و درمان کاهش می دهد. داده‌های پزشکی منبعی قدرتمند برای ایجاد تحلیل‌های ارزشمند و کاهش اتلاف داده‌ها هستند. با توجه به همه‌گیری جهانی ویروس کرونا و بروز چالش‌های بی شمار پزشکی، کلان‌داده می‌توانند به ارائه‌دهندگان خدمات بهداشتی در شناسایی الگوهای مربوط به سلامتی کمک شایانی کرده و ضمناً از این داده‌ها می‌توان برای اتخاذ اقدامات عملی در صنایع پزشکی استفاده کرد. گذشته از اینکه با استفاده از علم داده‌ها، تجربه بیماران از خدمات درمانی بهبود نسبتاً زیادی پیدا می‌کند، ذینفعان علاقه‌مند به بکارگیری کلان‌داده در صنعت پزشکی غالباً شامل ارائه دهندگان مراقبت‌های بهداشتی، صنعت فناوری سلامتی، داروسازها و بیمه‌های درمانی می‌باشند.

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 بیماری ها 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 می باشد، انجام گرفته است .افزایش قابل توجهی در داده های بهداشتی ، چه ساختاری و چه غیر ساختاری وجود دارد. سطح پیچیدگی بالای این داده ها مستلزم آن است که توسط تجزیه و تحلیل داده های بزرگ، پردازش شده و استنتاج های مربوط به آن بررسی شود و به صورت قابل اجرا ارائه گردد. Telemedicine همچنین در حال بهره برداری از قدرت داده های بزرگ برای بهبود امکانات مراقبت های بهداشتی موجود است. تاثیر فناوری اینترنت اشیاء و توسعه یادگیری ماشینی در مراقبت های بهداشتی امروز، تغییراز بیمارستان به خانه با آزمایش های پزشکی منظم و سایر خدمات بهداشتی است و همچنین استفاده از ابزارهای پزشکی را برای پزشکان و بیماران آسان تر می کند. به خصوص در موارد بحران، مراقبت های بهداشتی را برای بیماران ساده تر می کند. علاوه بر این، با انتقال فعالیت های عملی و اساسی به محیط های خانگی، بیمارستان ها می توانند بار کاری خود را کاهش دهند. کاهش هزینه ها نیز یکی از دستاوردهای اساسی این فناوری است[۴].

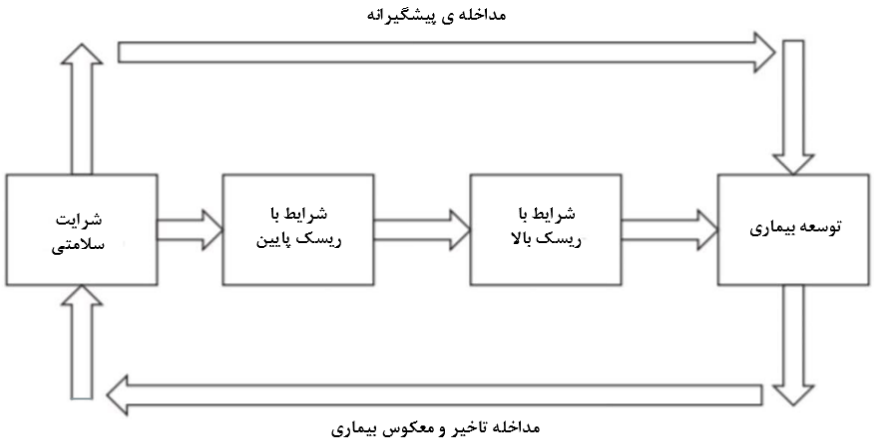
پژوهش[۱] رویکرد داده های بزرگ مربوط به مدیریت مراقبت های بهداشتی شخصی با استفاده از داده های بزرگ را مورد بحث قرار داده است. در این معماری نویسنده بر انسان فیزیولوژیکی مجازی (VPH) تمرکز کرده است که با پزشک از طریق دستگاهی کمکی ارتباط برقرار می کند. ضمن اینکه در این تحقیق در مورد پنج V کلان داده بحث شده و اهمیت پیاده سازی کلان داده در مراقبت های بهداشتی و اطلاعات شخصی توضیح داده شده است. در تحقیق [۲] با اجرای پرونده های الکترونیکی سلامت (EHR) که یک برنامه بهداشتی شخصی از طریق ساعت یا برنامه تلفن همراه است، بر سلامت بیمار نظارت می شود و پیاده سازی و نظارت با الزاماتی مانند اطلاعات بهداشت شخصی افراد انجام می گیرد. در سال ۲۰۰۹ استانداردهای بالینی توسط سازمان بهداشت جهانی (WHO) تغییر کرد و

اجرای فرآیندهای پزشکی با تغییرات زیادی در سیستم در مورد امنیت و دستکاری داده ها همراه شد. شکل ۲ پیاده سازی سوابق مراقبت های سلامت با تحلیل کلان داده را نشان می دهد.



شکل ۲: سوابق مراقبت های سلامت با تحلیل کلان داده

پژوهش [۳] با پیاده سازی اینترنت اشیاء در صنعت مراقبت های بهداشتی سروکار دارد. در این مدل، برخی از شناسایی های ریسک شامل ریسک کم خطر و متوسط و ریسک پرخطر نشان داده شده است. معماری ارائه شده قادر به پیش بینی بیماران در معرض خطر، کم خطر و یا بیماران در معرض ریسک بالا می باشد. شکل ۳ این معماری را نشان می دهد. در این معماری، رابطه ای بین وضعیت سلامتی ثبت شده با استفاده از دستگاه اینترنت اشیاء با توجه به الگوریتم های پیش بینی مورد نیاز برای درک کم خطر بودن یا بالا بودن آنها وجود دارد. به طوری که این معماری برای جلوگیری از مشکل و یا درمان بیمار مناسب می باشد.



شکل ۳: مدل مدیریت سلامت

مقاله [۵] به رویکردهای مختلف برای تحلیل پیش بینی می پردازد. سه نوع مختلف تجزیه و تحلیل داده ها وجود دارد. تجزیه و تحلیل توصیفی را می‌توان ساده‌ترین نوعِ تجزیه و تحلیل نامید. اندازه بزرگِ داده‌های بزرگ فراتر از درک انسان است و از این رو مرحله اول شامل خرد کردن داده‌ها و تبدیل آن‌ها به قطعات قابل درک است. هدف از این نوع تجزیه و تحلیل فقط جمع بندی یافته‌ها و درک آنچه در حال انجام است می‌باشد. تجزیه و تحلیل پیش‌بینی‌کننده شامل انواع تکنیک‌های آماری از داده‌کاوی ، مدل‌سازی پیش‌بینی‌کننده و یادگیری ماشینی است که واقعیت‌های کنونی و تاریخی را برای پیش‌بینی رویدادهای آینده یا در غیر این صورت ناشناخته تحلیل می‌کند. تجزیه و تحلیل تجویزی، که از تجزیه و تحلیل توصیفی و تجزیه و تحلیل پیش بینی کننده استفاده می‌کند و سپس بر اساس اطلاعات ارائه شده توسط این تکنیک‌های تحلیلی ، یک روش عملی را پیشنهاد می‌کند. در بسیاری از موارد ، تجزیه و تحلیل تجویزی هنوز در مراحل ابتدایی است و ابزارها و سیستم‌هایی که به تازگی منتشر شده‌اند می‌توانند اقدامات عملی را توصیه کنند. مقاله[۶] به شناسایی الگوهای فعالیت انسان با استفاده از داده های بزرگ سروکار دارد و این الگوها فرایند درک احساسات و رفتار انسان را برای شناسایی الگوهای سلامتی انسان تعریف می نماید.

۲.۵. سیستم موجود

سیستم های موجود بر پیاده سازی در حوزه ابری و مدیریت جریان داده لایو (زنده) تمرکز نمی کنند. مکانیسم های مختلفی مانند یادگیری ماشین با اسپارک و ماهوت برای ذخیره سازی داده ها و طراحی مدل پیش بینی در سرویس های ابری مورد توجه قرار می گیرند. اسپارک دارای دو کتابخانه برای یادگیری ماشین با APIهای مجزا اما الگوریتم های مشابه است. این مولفه‌های اسپارک توانایی استفاده مقیاس پذیر از الگوریتم‌های یادگیری را به توسعه‌دهندگان و تحلیلگران داده می‌دهد.پیش از اسپارک، معمولاً برای مدل‌های یادگیری ماشین در کلان‌داده از زبان‌های آماری مانندR، STATAو SAS استفاده می‌شد. سپس، مهندسان داده این مدل‌ها را مجدداً، برای مثال در جاوا، برای بکارگیری در هادوپ پیاده‌سازی کردند. اما، این نوع از جریان‌کار، فاقد بازده، مقیاس‌پذیری، توان‌پردازشی مناسب، و دقت بوده و زمان اجرا را طولانی می‌کند.با استفاده از اسپارک، همان مدل‌های یادگیری ماشین می‌توانند ساخته شوند و به کار گرفته شوند، در حالی که جریان‌کار را بسیار کارآمد، مقاوم و سریع‌تر می‌کند و در عمل کارایی بیشتر می‌شود. کتابخانه‌های یادگیری ماشین اسپارک در دو پکیج سپارک MLlib (spark.mllib) و اسپارک ML (spark.ml) .قرار دارند.

جریان داده لایو (زنده) که مشکل اصلی سیستم های موجود است تا کنون در مقالات مختلف مورد بحث قرار نگرفته است. همچنین جریان داده لایو (زنده) در حوزه پزشکی هنوز جهت پیاده سازی مورد توجه قرار نبوده است. در سیستم های موجود فعلی، مدیریت داده با اندازه و وزن ثابت وجود دارد، اما هیچ منبعی برای مدیریت جریان داده لایو (زنده) وجود ندارد. در ادامه این بخش، جریان داده لایو (زنده) در حوزه پزشکی مورد بحث قرار می گیرد.

۶. سیستم پیشنهادی

سیستم پیشنهادی دو نوع متفاوت را برای پیاده سازی به عنوان بهترین روش برای مدیریت مراقبت های بهداشتی با امنیت مناسب و بدون آسیب پذیری مورد بحث قرار می دهد. هر دو بخش روش پیشنهادی در ادامه ذکر شده است. این مقاله بر اساس سناریوهای ذیل طراحی شده است.

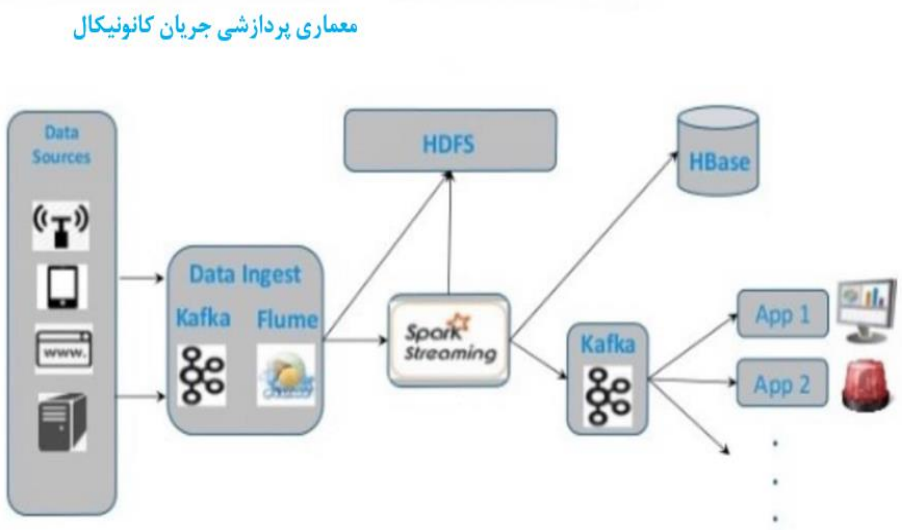
۷. بهترین رویکرد برای روش های فعلی

بهترین روش ها را نمی توان از ایجاد یک معماری تشخیص داد [۸-۷]. چرا که همه معماری ها برای یک نوع طراحی نشده اند. شکل ۲ بهترین نوع پیاده سازی است که تاکنون بدست آمده است. در این معماری، انواع مختلف مراقبت های بهداشتی و رویکردهایی برای حل مشکل نشان داده شده است [۱۱-۹]. تفکیک در فرآیند جمع آوری و پالایش داده ها وجود دارد و همچنین چارچوب و فرآیند جداگانه ای برای مکانیسم ذخیره سازی وجود دارد. در این معماری، مکانیزم ذخیره سازی، رویکردهای متفاوتی داشته است و از منابع ابری یا هر نوع مکانیزم ذخیره سازی اختصاصی دیگری استفاده شده است .در این معماری، MapReduce یک مدل پردازش داده با مقیاس پذیری

اسان پردازش داده روی چندین گره محاسباتی است این پارادایم برنامه ها را دردو فاز نگاشت و کاهش اجرا می کند هرفاز توسط یک تابع تعریف می شود که توابع Mapper Reducer نام دارند.

۸. رویکرد پیشنهادی

آپاچی اسپارک به عنوان چارچوب کلان داده برای پیاده سازی داده های لایو (زنده) استفاده می شود. پردازش داده های لایو (زنده) مهمترین مسائل در کلان داده است و این می تواند پیاده سازی بسیار مفیدی برای متدولوژی موجود باشد. آپاچی اسپارک یک فریم ورک محاسبات توزیع شده همه منظوره اوپن سورس است. آپاچی اسپارک واسط برای برنامه نویسی کل توزیع ها با موازی سازی داده ها و تلرانس را فراهم می کند. آپاچی اسپارک به عنوان پایه معماری خود، مجموعه داده های توزیع شده انعطاف پذیر (RDD)، یک مجموعه داده فقط خواندنی از داده های توزیع شده در یک توزیع از کلاستر ماشین ها، که به شکل تلرانس حفظ می شود را قرار داده است. کافکا بهترین روش برای مدیریت پیاده سازی بهتر داده های جریان لایو است [۱۳-۱۲]. داده های بورس اوراق بهادار و داده های پزشکی مختلف مثالی از داده های جریان لایو (زنده) می باشد. به عنوان مثال، داده های پرواز می توانند داده های لایو (زنده ای) باشند که برای مدیریت پروازی مورد نیاز است. [۱۵-۱۴]. استریم اسپارک در شکل ۵ استریم توضیح داده شده است.



شکل ۴: جریان اسپارک برای مدیریت داده های لایو (زنده)

اگرچه مدیریت پیاده سازی بهتر داده های جریان لایو روی آپاچی اسپارک نیز امکان پذیر هست، اما کافکا و فلوم برای مکانیسم جذب داده ها نیز قابل اعتماد هستند [۱۶]. به عبارت دقیق‌تر، اگر بیماری مشکلی داشته باشد و امکان درخواست کمک یا اطلاع‌رسانی به پزشک متخصص وجود نداشته باشد، با استفاده از الگوریتم یادگیری ماشین، یک دستگاه اینترنت اشیاء که به فضای ابری متصل است، این کار را انجام خواهد داد. جزئیات مربوط به سلامت بیمار و اطلاعاتی که به عنوان ورودی در نظر گرفته شده بود، توسط مدل های یادگیری ماشین به عنوان داده واقعی یا داده غیر واقعی تأیید می شود [۱۸-۱۷]. اگر داده ها واقعی باشد، اطلاعات پردازش می شود و پزشک اطلاعات بیمار را دریافت می کند. در این سیستم، داده ها، داده های لایو (زنده) هستند. با استفاده از داده های لایو (زنده) و دستگاهی که در آن مدل های یادگیری ماشینی تعبیه شده است، بر اساس یک سری مکانیزم های داخلی، داده های لایو (زنده) ممیزی شده و مدیریت می شوند. در این معماری، اسکالا به عنوان فریم ورک اصلی و کافکا به عنوان دریافت کننده داده از یک پلتفرم کلان داده حضور دارند.

۹. نتیجه گیری

کلان داده مهم ترین مکانیزمی است که در سال های اخیر برای مدیریت داده ها به اشکال مختلف طراحی شده است. تمام اطلاعات جمع‌آوری شده از منابع مختلف با استفاده از مکانیزم‌های داده کاوی پیش پردازش شده و ویژگی‌های مهم داده ای با توجه به نوع مساله و نحوه پیاده سازی به فریم ورک ارسال می‌شوند. در این پژوهش، موضوع پیاده سازی پلتفرم های مختلف کلان داده در حوزه پزشکی مورد بررسی قرار گرفت. در این پلتفرم، اطلاعات جمع‌آوری‌شده از منابع مختلف تابع مدل‌های پیش‌بینی مبتنی بر یادگیری ماشین بوده و داده های جمع آوری شده بایستی واقعی بوده و در آن داده پرت یا گم شده وجود نداشته باشد. اگر ویژگی داده ای خاصی بر روی این پلتفرم کلان داده با استفاده از هر نوع ذخیره سازی ابری ایجاد شده باشد، در مرحله بعدی بایستی در پایگاه داده نگاشت شود. این مدل برنامه نویسی بطور موفقیت آمیزی در بسیاری از برنامه های گوگل (نظیر سیستم های پردازش متون و جستجوی محتوای وب) برای پردازش حجم عظیم داده های وب که بروی سیستم فایل توزیع شده GFS و BigTable قرار دارند، مورد استفاده قرار گرفته است.

منابع

[۱] Viceconti, M., Hunter, P., & Hose, R. (۲۰۱۵). Big data, big knowledge: big data for personalized healthcare. IEEE journal of biomedical and health informatics, ۱۹(٤), ۱۲۰۹-۱۲۱۵.
[۲] Zhang, C., Ma, R., Sun, S., Li, Y., Wang, Y., & Yan, Z. (۲۰۱۹). Optimizing the electronic health records through big data analytics: a knowledge-based view. IEEE Access, ۷, ۱۳٦۲۲۳-۱۳٦۲۳۱.
[۳] Ma, Y., Wang, Y., Yang, J., Miao, Y., & Li, W. (۲۰۱٦). Big health application system based on health internet of things and big data. IEEE Access, ٥, ۷۸۸٥-۷۸۹۷.
[٤] Fizi, F., & Askar, S. (۲۰۱٦). A novel load balancing algorithm for software defined network-based datacenters, International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom), Graz, ۲۰۱٦, pp. ۱-٦
[٥] Kumar, S., & Singh, M. (۲۰۱۸). Big data analytics for healthcare industry: impact, applications, and tools. Big data mining and analytics, ۲(۱), ٤۸-٥۷.
[٦] Yassine, A., Singh, S., & Alamri, A. (۲۰۱۷). Mining human activity patterns from smart home big data for health care applications. IEEE Access, ٥, ۱۳۱۳۱-۱۳۱٤۱.
[۷] Sun, H., Liu, Z., Wang, G., Lian, W., & Ma, J. (۲۰۱۹). Intelligent analysis of medical big data based on deep learning. IEEE Access, ۷, ۱٤۲۰۲۲-۱٤۲۰۳۷.
[۸] Jiang, C., & Li, Y. (۲۰۱۹). Health big data classification using improved radial basis function neural network and nearest neighbor propagation algorithm. IEEE Access, ۷, ۱۷٦۷۸۲-۱۷٦۷۸۹
[۹] Zhang, J., Zhao, T., & Zhu, P. (۲۰۱۹). Analysis method of motion information driven by medical big data. IEEE Access, ۷, ۱۷٤۱۸۹-۱۷٤۱۹۹.
[۱۰] Harerimana, G., Jang, B., Kim, J. W., & Park, H. K. (۲۰۱۸). Health big data analytics: A technology survey. IEEE Access, ٦, ٦٥٦٦۱-٦٥٦۷۸.
[۱۱] Nazir, S., Khan, M. N., Anwar, S., Adnan, A., Asadi, S., Shahzad, S., & Ali, S. (۲۰۱۹). Big data visualization in cardiology—a systematic review and future directions. IEEE Access, ۷, ۱۱٥۹٤٥-۱۱٥۹٥۸.
[۱۲] Dhayne, H., Haque, R., Kilany, R., & Taher, Y. (۲۰۱۹). In search of big medical data integration solutions-a comprehensive survey. IEEE Access, ۷, ۹۱۲٦٥-۹۱۲۹۰.
[۱۳] Panayides, A. S., Pattichis, M. S., Leandrou, S., Pitris, C., Constantinidou, A., & Pattichis, C. S. (۲۰۱۸). Radiogenomics for precision medicine with a big data analytics perspective. IEEE journal of biomedical and health informatics, ۲۳(٥), ...۲۰٦۳-۲۰۷۹

[۱۴] Sahoo, P. K., Mohapatra, S. K., & Wu, S. L. (۲۰۱٦). Analyzing healthcare big data with prediction for future health condition. IEEE Access, ٤, ۹۷۸٦-۹۷۹۹.

[۱۵] Carnimeo, L., Trotta, G. F., Brunetti, A., Cascarano, G. D., Buongiorno, D., Loconsole, C., ... & Bevilacqua, V. (۲۰۱۹). Proposal of a health care network based on big data analytics for pds. The Journal of Engineering, ۲۰۱۹(۶), ۴۶۱۱-۴۶۰۳.

[۱۶] Hu, H., Wen, Y., Chua, T. S., & Li, X. (۲۰۱٤). Toward scalable systems for big data analytics: A technology tutorial. IEEE access, ۲, ٦٥۲-٦۸۷.

[۱۷] Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (۲۰۱٥). The internet of things for health care: a comprehensive survey. IEEE access, ۳, ۷۰۸-۶۷۸.

[۱۸] Tsang, G., Xie, X., & Zhou, S. M. (۲۰۱۹). Harnessing the power of machine learning in dementia informatics research: Issues, opportunities, and challenges. IEEE reviews in biomedical engineering, ۱۳, ۱۱۳-۱۲۹.