

## سامانه سنها و تحلیل داده‌های آن بر اساس رویکرد SMART

نجمه ملایی<sup>۱</sup>

عباس طهماسبی<sup>۲</sup>

### چکیده

هم‌اکنون در صنعت بیمه، سامانه‌ی سنها تنها منبع داده محسوب می‌شود که با تکیه بر آن نظارت، کنترل و شفافیت صنعت بیمه بیشتر شده و نهادهایی که نیاز به اطلاعات صنعت بیمه دارند می‌توانند از این اطلاعات استفاده کنند. ایجاد زیرساخت‌های لازم به‌منظور استفاده از این داده‌ها، مستلزم ارزیابی، انتخاب، مهار فناوری و همچنین ترکیب، یکپارچه‌سازی، تولید و بومی‌سازی فناوری‌هاست. الزامات نهفتگی در داده‌های سنها به‌صورت دسته‌ای و داده‌ها به‌صورت ساختاریافته می‌باشند. نهاد ناظر صنعت بیمه؛ بیمه مرکزی ج.ا.ایران، متولی مدیریت و ذخیره‌سازی و امنیت داده‌های سنها است. در صورتی‌که کیفیت داده‌ها را کامل فرض کنیم انتخاب استراتژی داده محور و الگوهای تحلیل مناسب داده‌های سنها می‌تواند ارزش‌آفرین باشد. در این مقاله مراحل تحلیل داده در سامانه سنها مطابق رویکرد SMART بررسی و تبیین شده است.

واژگان کلیدی: سامانه سنها، تحلیل داده، رویکرد SMART

## مقدمه

در گذشته نه چندان دور، شرکت‌های بیمه در پایان هر ماه انبوهی از بردروهای صدور و خسارت خود را بر روی کاغذ چاپ کرده و جهت بررسی به بیمه مرکزی ارسال می‌کردند. بعدها بیمه مرکزی به جای پرینت‌های کاغذی، بردروها را در قالب فایل اکسل از شرکت‌های بیمه دریافت می‌نمود. تا در اواسط سال ۱۳۹۰ سامانه سنهاب بیمه مرکزی ایران به وجود آمد و از طریق این سامانه شرکت‌ها نیازی به ارسال بردرو ندارند و بیمه مرکزی اطلاعات بردروها را از بانک‌های اطلاعاتی شرکت‌ها به صورت مستقیم استخراج نموده و با استفاده از آنها صورت‌حساب‌های اتکایی اجباری را به صورت خودکار صادر نموده و جهت تایید به شرکت‌های بیمه ارائه می‌نماید. این روش کار با آن‌که به نسبت قبل پیشرفت چشمگیری محسوب می‌شود اما هنوز اشکالاتی متوجه آن می‌باشد که از آن جمله وابستگی زیاد به بانک اطلاعاتی متفاوت شرکت‌ها می‌باشد که بیمه مرکزی را با پیچیدگی‌های فراوان جهت استخراج اطلاعات لازم صورت‌حساب‌ها مواجه کرده و از طرف دیگر همچنان یک روش آفلاین می‌باشد (قلاتی و خاکور، ۱۳۹۴). لذا می‌توان گفت الزامات نهفتگی داده‌های سنهاب از نوع دسته‌ای می‌باشند. همچنین فرمت پایگاه‌های داده آن در حال حاضر SQL است.

اگرچه کیفیت داده معمولاً به مناسب بودن داده برای استفاده اطلاق می‌شود (جوران، ۱۹۷۴). "مناسب برای استفاده" یک مفهوم چند بعدی است که هر دوی مفهوم ذهنی و معیار عینی براساس مجموعه داده‌های مورد درخواست را داراست (پیبینو و همکاران، ۲۰۰۲). بی‌کیفیت شدن داده‌ها علل گوناگونی مانند خطای انسانی، فقدان منابع داخلی، استراتژی نامناسب، عدم کفایت تکنولوژی مرتبط در حال استفاده، فقدان تکنولوژی مرتبط، فقدان مهارت مورد نیاز برای استفاده صحیح از تکنولوژی موجود، فقدان ارتباطات داخلی بین واحدها، ناکافی بودن حمایت مدیر ارشد و بودجه ناکافی دارد. برای کسب و کارها تبدیل داده‌ها به دانش مطلوبیت دارد، اگر داده‌ها کیفیت بالایی داشته باشند این مطلوبیت افزایش می‌یابد (گزارش تطبیق مدیریت داده جهانی، ۲۰۱۶).

با کامل شدن فناوری اطلاعات و تکنولوژی‌های مربوط به آن سازمان‌ها باید از این اهرم به عنوان ارتباطات و تراکنش‌های امن و دقیق با مشتریان و تأمین‌کنندگان و ذینفعان استفاده کنند (پاترسون، ۲۰۱۵). هم‌اکنون بسیاری از سازمان‌ها به آهستگی برای اتخاذ فناوری‌های داده عمل می‌کنند (کاسترو و کورت، ۲۰۱۳). ذخیره‌سازی و تحلیل داده‌ها از فناوری‌های داده محسوب می‌شوند. تحلیل داده<sup>۱</sup> فرآیند بررسی داده‌ها برای پیدا کردن حقایق، روابط، الگوها، بینش‌ها یا روندهاست. هدف کلی از تحلیل داده پشتیبانی از تصمیم‌گیری بهتر است. انجام تحلیل داده به ایجاد الگوها و روابط میان داده‌های در حال تحلیل کمک می‌کند (ارل و همکاران، ۲۰۱۵).

از جمله قابلیت‌ها و پتانسیل‌های سامانه سنهاب می‌توان به موارد زیر اشاره نمود:

- توسعه سامانه سنهاب برای انجام نظارت مالی دقیق در مورد شرکت‌های بیمه
- توسعه سامانه سنهاب برای ارائه انواع اعلام‌های آنلاین در مورد سوابق صدور و خسارت بیمه‌گذاران، بیمه‌شده‌ها و خودروها به شرکت‌های بیمه متقاضی
- تحلیل داده‌های جمع‌آوری شده در سامانه سنهاب برای استخراج شاخص‌ها و داشبوردهای دقیق عملکردی صنعت بیمه جهت استفاده در بیمه مرکزی و شرکت‌ها بیمه

- توسعه سامانه سنهاب در مورد انواع عملیات اتکایی قبولی و واگذاری اختیاری و ارائه انواع اعلام‌های لازم به شرکت‌های بیمه
- توسعه سامانه سنهاب در مورد ارائه سرویس به شرکت‌های بیمه در مورد بیمه‌نامه‌های کنسرسیومی (قلاتی و خاکور، ۱۳۹۴).

در این پژوهش به بررسی قابلیت سوم یعنی «تحلیل داده‌های جمع‌آوری شده در سامانه سنهاب» به منظور تصمیم‌گیری بهتر و ایجاد دانش مطلوب از آن براساس رویکرد SMART پرداخته شده است.

## ۱- رویکرد SMART

اکثر سازمان‌ها بررسی می‌کنند چگونه ظرفیت‌شان را برای کاوش و تحلیل افزایش دهند درحالی‌که ممکن است در داده‌ها غرق شوند و نتوانند بینش‌های واقعی را استخراج کنند. همان‌طورکه در نمودار ۱ مشاهده می‌شود، رویکرد SMART به‌طور موثرتری جمع‌آوری و تحلیل داده‌ها را به‌صورت ویژه<sup>۱</sup>، قابل اندازه‌گیری<sup>۲</sup>، عملی<sup>۳</sup>، واقعی<sup>۴</sup> و به‌موقع<sup>۵</sup> فیلتر می‌کند (SMART). برخلاف رویکرد سنتی در این رویکرد به جای جمع‌آوری داده در گام اول استراتژی تحلیل داده مشخص می‌شود. این استراتژی داده‌محور است که تصمیمات را از قبل تعیین می‌کند. گام‌های رویکرد هوشمند شامل مشخص کردن بینش لازم، مشخص کردن داده‌های ضروری برای چنین بینشی، جمع‌آوری و مدیریت داده‌های ضروری به‌منظور کسب بینش است درحالی‌که در رویکرد سنتی بعد از جمع‌آوری و تحلیل داده تصمیمات تعیین می‌شوند و بعد از آن داده‌ها به انبار منتقل می‌شوند.

نمودار ۱. رویکرد هوشمند در مقابل رویکرد سنتی برای تحلیل داده‌ها<sup>۶</sup>



## ۲-۱- تصمیمات از قبل تعیین شده

دلایلی مانند افزایش کارایی، بهبود رضایت مشتری، توانایی تصمیم‌گیری آگاهانه‌تر، کاهش هزینه‌ها، حفظ اعتبار سازمان‌ها، افزایش نوآوری در داده، کشف فرصت‌های بازار از طریق پروفایل مشتری و کاهش ریسک تقلب سازمان‌ها را وامی‌دارد تا برای ایجاد استراتژی داده‌های باکیفیت برنامه‌ریزی کنند (گزارش تطبیق مدیریت داده جهانی، ۲۰۱۶). ۶ کاربرد از موارد یادشده را می‌توان در جدول ۱ مشاهده نمود.

1. Specific
2. Measurable
3. Actionable
4. Realistic
5. Timely
6. PwC Insurance 2020: The digital prize – Taking customer connection to a new level

جدول ۱. نمونه‌هایی از کاربرد داده‌ی بزرگ (روحانی و حسینی، ۱۳۹۴)

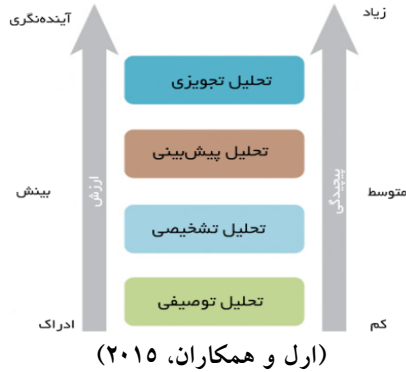
کاربرد	مشخصات	منابع داده
مانیتورینگ و بهینه‌سازی شبکه‌ی انرژی	گلوگاه داده	داده‌های حسگر حاصل از کنتورهای هوشمند
	گلوگاه محاسبات	
	حجم زیاد داده	
تشخیص تقلب‌ها در کارت‌های اعتباری	گلوگاه داده	داده‌های نقطه‌ی فروش یا کارت‌خوان‌ها (POS)
	گلوگاه محاسبات	پروفایل مشتریان/اطلاعات مشتریان
	حجم زیاد داده	سوابق تراکنش‌ها
	موازی سازی	مدل‌ها و الگوهای پیش‌بینی شده
	تنوع و گوناگونی داده‌ها	
پروفایل اطلاعات	حجم زیاد داده‌های موازی	منابع منتخب برای پیشنهادهای تجاری
خوشه‌بندی و طبقه‌بندی مشتریان	گلوگاه داده	پروفایل مشتریان/اطلاعات مشتریان
	گلوگاه محاسبات	سوابق تراکنش‌ها
	حجم زیاد داده	مجموعه داده‌های افزایش یافته/ارتقاء یافته
	موازی سازی	
	تنوع و گوناگونی داده‌ها	
موتورهای توصیه‌گر	گلوگاه داده	پروفایل مشتریان/اطلاعات مشتریان
	گلوگاه محاسبات	سوابق تراکنش‌ها
	حجم زیاد داده	مجموعه داده‌های افزایش یافته/ارتقاء یافته
	موازی سازی	داده‌های مربوط به شبکه‌های اجتماعی
	تنوع و گوناگونی داده‌ها	
مدل‌سازی قیمت	گلوگاه داده	داده‌های نقطه‌ی فروش یا کارت‌خوان‌ها (POS)
	گلوگاه محاسبات	پروفایل مشتریان/اطلاعات مشتریان
	حجم زیاد داده	سوابق تراکنش‌ها
	موازی سازی	مدل‌ها و الگوهای پیش‌بینی شده

در این بخش «اهداف و موانع کسب‌وکار» مشخص می‌شوند. مثلاً در شرکت‌های بیمه‌ای فرار مشتری به علت کاهش سرعت پردازش خسارت‌ها یا ظهور رقبای فناوری محور باشد. در این مرحله کمیته‌ای از مدیران ارشد سازمان به‌منظور بررسی مشکل ایجاد می‌شود. تنظیم حق‌بیمه نادرست، عدم تطابق سریع و چابک با مقررات جدید، عدم پیاده‌سازی صحیح سیستم مدیریت ریسک و ... باشد. (ارل و همکاران، ۲۰۱۵). تصمیمات از قبل تعیین‌شده با یکی از رویکردهای تحلیل داده توصیفی، تشخیصی، پیش‌بینانه و تجویزی قابل تحلیل هستند. رویکرد تحلیلی داده‌ها، رشته‌ای شامل مدیریت دوره عمر کامل داده‌هاست که در آن جمع‌آوری، پاک‌سازی، سازماندهی، نگهداری، تحلیل و اداره داده‌ها انجام می‌شود. انواع رویکرد تحلیلی داده، شرایط تصمیم‌گیری داده‌محور با پشتوانه علمی را فراهم می‌کند به‌طوری‌که تصمیمات می‌توانند مبتنی بر داده‌های واقعی شکل گیرند، نه صرفاً براساس تجربه‌های گذشته یا شواهد. چهار دسته کلی از رویکرد تحلیلی وجود دارد که توسط نتایج به‌دست آمده متمایز می‌شوند:

- رویکرد تحلیلی توصیفی<sup>۱</sup>: رویکرد تحلیلی توصیفی اغلب از طریق گزارش موقت یا داشبوردها انجام می‌شود.
- رویکرد تحلیلی تشخیصی<sup>۲</sup>: نتایج رویکرد تحلیلی تشخیصی از طریق ابزارهای مصورسازی تعاملی مشاهده می‌شوند که کاربران را قادر به شناسایی روندها و الگوها می‌کند.

- رویکرد تحلیلی پیش‌بینانه<sup>۱</sup>: پیش‌بینی‌ها براساس الگوها، روندها و استثنائات موجود در داده‌های تاریخی و جاری ایجاد می‌شوند.
  - رویکرد تحلیلی تجویزی<sup>۲</sup>: رویکرد تحلیلی تجویزی ارزش بیشتری از هر نوع دیگری از رویکرد تحلیلی ارائه می‌کند و به تناسب، نیاز به پیشرفته‌ترین مهارت‌ها و هم‌چنین نرم‌افزارها و ابزارهای تخصصی دارد.
- انواع رویکردهای تحلیلی مختلف از تکنیک‌ها و الگوریتم‌های رویکرد تحلیلی مختلف استفاده می‌کنند. این به معنی آن است که ممکن است الزامات پردازش و ذخیره‌سازی داده برای تسهیل ارائه انواع مختلفی از نتایج تحلیلی تغییر کنند. شکل ۱ این واقعیت را نشان می‌دهد که تولید نتایج تحلیلی با ارزش، پیچیدگی و هزینه محیط تحلیلی را افزایش می‌دهد.

شکل ۱. ارزش و پیچیدگی در هر یک از رویکردهای تحلیلی



## ۲-۲- مشخص کردن بینش لازم

حوزه ویژه‌ای که در آن داده‌ها به‌وجود می‌آیند، انواعی از معماری را مشخص می‌کنند که برای ذخیره‌سازی، پردازش و انجام تحلیل‌ها روی آن نیاز است (گروه کاری داده‌های بزرگ، ۲۰۱۴). سپس با توجه به نوع و مشخصات داده‌های بزرگ ذخیره‌شده در صنعت مربوطه رویکردهای تحلیلی مناسب و به تبع آن نرم‌افزارهای مناسب بررسی و پیاده‌سازی می‌شود. سؤالات ذیل جهت مشخص کردن بینش لازم و سنجش میزان پذیرش و آمادگی سازمان باید پاسخ داده شوند. این سؤالات میزان آمادگی سازمان و درواقع فرهنگ سازمان از نظر امکان‌سنجی، توجیه‌پذیری و معقولیت داده، ارزش، قابلیت یکپارچگی و نگهداری مستمر و توسعه پایدار است.

- سؤال ۱ (امکان سنجی): آیا نهاد ناظر (بیمه مرکزی) تکنولوژی‌های جدید و نوظهور را مورد پذیرش، بررسی و آزمایش قرار می‌دهد؟
- سؤال ۲ (امکان سنجی): آیا نهاد ناظر (بیمه مرکزی) منابع سازمانی خود را برای رویارویی با چالش‌های پیش‌رو در حوزه تکنولوژی داده‌ی بزرگ در نظر می‌گیرد؟
- سؤال ۳ (توجیه‌پذیری و معقولیت داده): آیا نهاد ناظر (بیمه مرکزی) رویکرد داده محور<sup>۳</sup> (محوریت داده) و استفاده از منابع داده‌ای بیشتر جهت تحلیل و گزارش‌دهی را در آینده در دستور کار قرار خواهد داد؟
- سؤال ۴ (ارزش): آیا شاخص‌ها و روش‌های مشخصی برای اندازه‌گیری دستاوردهای حاصل از به‌کارگیری و پیاده‌سازی داده‌ی بزرگ تعیین و تعریف شده است؟
- سؤال ۵ (قابلیت یکپارچگی): چه مراحل و گام‌هایی برای سنجش یکپارچگی سنها با سایر اجزا در پژوهشکده وجود دارد که باید طی شود؟

1. Predictive Analytics  
2. Prescriptive Analytics  
3. Data-Centric

- سؤال ۶ (نگهداری مستمر و توسعه پایدار): چه برنامه‌ای برای شناسایی دقیق محدودیت‌ها و موانع به‌منظور مدیریت مستمر و نگهداری پایدار جهت پیاده‌سازی هرچه بهتر داده‌ی بزرگ می‌توان تدوین کرد؟ (روحانی و همکاران، ۹۴).

### ۲-۳- مشخص کردن داده‌های لازم برای چنین بینشی

در این مرحله داده‌های مرتبط و مشخصه‌های مربوطه متناسب با تصمیم از قبل اتخاذشده، معرفی می‌گردد. به‌عنوان مثال در صورتی که هدف از تحلیل شناسایی خسارت‌های تقلب باشد فیلدهای نمایش داده‌شده در جدول ۲ می‌تواند اثرگذار باشد و از داده‌های دیگر باید تفکیک شوند.

جدول ۲. عناوین فیلدها

نام	سن	سن مقصر	جنسیت	جنسیت	منطقه	نوع	نوع	سن	تاخیر در اعلام	تاریخ
فیلد	زیان‌دیده		زیان‌دیده	مقصر	خسارت	خسارت	خودروی مقصر	خودروی مقصر	خسارت مقصر یا زیان‌دیده	وقوع خسارت
نام لاتین	Injured or Dead Age	Criminal Age	Injured or Dead Gender	Criminal Gender	Loss Region	Type of Loss	Type of Vehicle	Age of Vehicle	Delay at Annunciation	Date of Loss

مثالی دیگر از داده‌های مربوط به شناسایی صدمات غیرمرتبط با حادثه در میان سایر صدمات است. هر خسارت بدنی شامل مجموعه‌ای از داده‌های دیه مقدر و ارش است. در صورت جمع‌آوری فیلدهای نمایش داده‌شده در شکل ۲ تا حدودی می‌توان صدمات نامربوط و غیرمتعارف را شناسایی کرد. در این مثال با توجه به اینکه تمامی صدمات مربوط به دست و پا در حد بسیار جزئی است لذا وجود صدمه گوش چپ غیرعادی است.

شکل ۲. آسیب‌های ثبت‌شده در یک سیستم خسارت به فرم داده ساختار یافته

ردیف	نام و نوع اندام صدمه دیده	دیات	ارش	جمع درصد خسارت	شرح نظریه پزشکی	مغایرت
	دست/پا--کیودی	0.1500	0.0000	0.1500	زانوی راست	0.0000
	دست/پا--حارصه	0.5000	0.0000	0.5000	ساعد راست	0.0000
	دست/پا--حارصه	0.5000	0.0000	0.5000	آرنج راست	0.0000
	دست/پا--دامیه	1.0000	0.0000	1.0000	بشت ساق چپ	0.0000
	دست/پا--کیودی	0.1500	0.0000	0.1500	ساق پای چپ	0.0000
	گوش--از بین رفتن حس شنوایی هر گوش	50.0000	0.0000	50.0000	گوش چپ	0.0000
	ارش تورم	0.0000	0.5000	0.5000	ساق پای چپ	0.0000

### ۲-۴- جمع‌آوری و مدیریت داده‌های ضروری

حجم‌های بزرگ از داده با سرعتی بسیار سریع‌تر در انواع فرمت‌ها مانند رسانه‌ها و متن تولید می‌شوند که به راحتی با ساختار پایگاه داده ستون و ردیف تطابق ندارند. با در نظر گرفتن این فاکتورها راه‌حل‌های زیادی ابداع شده که علاوه بر فراهم کردن مقیاس، نیاز توسعه‌دهندگان را نیز در زمان ساخت برنامه‌های کاربردی اجتماعی، تحلیلی، بازی، مالی یا پزشکی با مجموعه داده‌های بزرگ سرعت می‌بخشند<sup>۱</sup>. نمودار ۲ طبقه‌بندی برای انواع متفاوت پایگاه داده را نشان می‌دهد که برای ذخیره‌گاه داده‌ها استفاده شده است. رایج‌ترین مدل‌های پایگاه داده New SQL، گراف، Daynomo -القاشده، جدول بزرگ هوشمند، کلید-مقدار، مستند و رابطه‌ای<sup>۲</sup> است. الزامات نهفتگی اولین روش برای طبقه‌بندی داده‌ها براساس گستره زمانی‌ای است که طبق آن داده‌ها باید تحلیل شوند: بلادرنگ (جریان‌های مالی، پردازش رویداد پیچیده (CEP)، تشخیص مزاحمت،

1. <http://blog.softlayer.com/2012/breaking-down-big-data-database-models/>

2. "Breaking Down 'Big Data' - Database Models - SoftLayer Blog." <http://blog.softlayer.com/2012/breaking-down-big-data-database-models/>.

شناسایی تقلب)، نزدیک به بلادرنگ (آگهی دادن)، دسته‌ای (خرده‌فروشی، پزشکی قانونی، بیوانفورماتیک، داده‌های جغرافیایی، داده‌های تاریخی با انواع مختلف) (گروه کاری داده بزرگ، ۲۰۱۴). الزامات نهفتگی داده‌های سنها از نوع دسته‌ای و داده‌ها به صورت ساختاریافته می‌باشند. همچنین فرمت پایگاه‌های داده آن در حال حاضر SQL است. با توجه به ساختاریافته بودن داده‌های سنها به صورت سطر و ستون و مشخص بودن عناوین فیلدها به راحتی می‌توان داده‌های مدنظر را برای شناسایی پینش تعیین شده مشخص و انتخاب کرد.

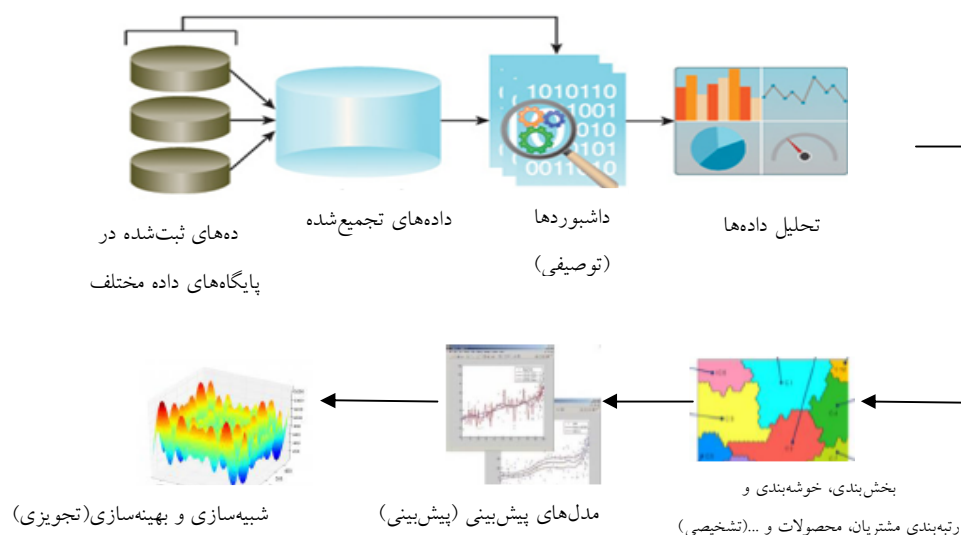
## نمودار ۲. انواع منابع ذخیره‌گاه داده‌ها



(گروه کاری داده بزرگ، ۲۰۱۴)

الگوهای تحلیل به چندجمله‌ای، MARS، درخت‌های تصمیم، شبکه‌های بیزین، ماشین‌های بردار پشتیبان، میانگین  $k$ ، ترکیب‌های گوسین، قاعده تحلیل جزء، فعال، هم‌آموزی، فرآیندهای تصمیم‌گیری مارکوف و یادگیری-Q تقسیم می‌شود. زیرساخت ذخیره‌گاه نیز شامل رابطه‌ای، مستند، کلید-مقدار، جدول بزرگ هوشمند، Dynamo القاشده، گراف و New SQL است. (گروه کاری داده بزرگ، ۲۰۱۴) هر یک از الگوهای تحلیل براساس تصمیم اخذشده و بر مبنای یکی از رویکردهای تحلیل توصیفی، تشخیصی، پیش‌بینانه و تجویزی جهت تحلیل داده‌ها انتخاب می‌شوند. در شکل ۳ این واقعیت در فرآیند کسب‌وکار نمایش داده شده است. دوره عمر رویکرد تحلیلی داده‌ها شامل، ارزیابی مورد کسب‌وکار، شناسایی مشخصات داده‌ها، شناسایی انواع داده‌ها، اکتساب و پالایش داده‌ها، استخراج داده‌ها، اعتبارسنجی داده‌ها و پاک‌سازی داده‌ها، جمع و بازنمایی داده‌ها، تحلیل داده‌ها، مصورسازی داده‌ها و استفاده از نتایج تحلیل است (ارل و همکاران، ۲۰۱۵).

### شکل ۳. انواع رویکردهای تحلیل در فرآیند کسب و کار به ترتیب ارزش (value)



(دلوتی، ۲۰۱۴)

با استفاده از خوشه‌بندی در داده‌کاوی، می‌توان یک جمعیت نامنظم را به مجموعه‌ای از زیرگروه‌های منظم تقسیم‌بندی نمود. اشیاء براساس اصل ماکزیم شباهت بین اعضاء هر خوشه و مینیم شباهت بین خوشه‌های مختلف، گروه‌بندی می‌شوند. یعنی خوشه‌ها به‌گونه‌ای تنظیم می‌شوند که اشیاء داخل هر خوشه بیشترین شباهت را با یکدیگر داشته باشند. معیار شباهت وقتی که همه مشخصه‌ها پیوسته هستند، معمولاً با فاصله اقلیدسی بیان می‌شود و در غیر این صورت یک معیار مناسب برای آن در نظر گرفته می‌شود (هان و همکاران، ۲۰۰۶). از جمله تکنیک‌های مورد استفاده عبارتند از: الگوریتم K-means و روش اعتبارسنجی سیلوئت. از طرفی تکنیک‌های یادگیری ماشین که زیرمجموعه روش‌های داده‌کاوی‌اند، روش‌های خودکار و مقیاس‌پذیری را ممکن می‌سازد تا بینش‌ها از داده‌های بزرگ چند بعدی بتواند تراوش کند. به‌طور گسترده یادگیری ماشین توانایی کامپیوترها به یادگیری خودکار الگوها و استخراج از داده‌ها است. این الگوریتم‌ها می‌توانند در راستای محورهای متفاوت زیادی دسته‌بندی شوند.

### ۳- نتایج و پیشنهادها

سنهاب لینک شرکت‌های بیمه با بیمه مرکزی است. لینک سنهاب اکنون یکی از منابع داده در بیمه مرکزی است که شرکت‌های بیمه را به یکدیگر و بیمه مرکزی مرتبط ساخته است. در صورت به‌روزرسانی اطلاعات آن به‌صورت بلادرنگ، جهت یکپارچه‌سازی برنامه‌های کاربردی در تمامی شرکت‌های بیمه، این سامانه می‌تواند به‌عنوان منبع عظیمی از داده‌ها، نقشی کلیدی برای اجرای قوانین جدید از طریق ارتباط ارگان‌های درگیر در فرآیند خسارت مورد استفاده قرار می‌گیرد که در نهایت خود منجر به ایجاد پایگاه عظیمی از داده‌های خسارت خواهد بود که در تحلیل‌ها، راهکارها و پیشنهادهای آتی مورد استفاده قرار خواهد گرفت.

کاربردی که هم‌اکنون سنهاب در بیمه مرکزی و برای شرکت‌های بیمه دارد سیستم اعلام براساس فیلدهای مختلفی مانند کد ملی، شماره موتور، شماره شاسی و ... است که این کاربرد تنها رویکرد توصیفی از این پایگاه داده عظیم می‌باشد با



ایجاد کمیته‌ای جهت انجام تحلیل‌های داده در بیمه مرکزی می‌توان علاوه بر این کاربرد از رویکردهای تحلیل تشخیصی مانند تشخیص تقلبات، تجویزی مانند قیمت‌گذاری و پیش‌بینی مانند پیش‌بینی میزان خسارت به‌وقوع پیوسته در سال‌های آتی بهره‌جست. با توجه به مجموعه داده‌ها در سنهاب در چرخه زندگی یک بیمه‌نامه (صدور، خسارت و تمدید) و یا طول عمر مشتری (پروفایل‌ها، سرشماری، شکایت و...) با انتخاب یک رویکرد استراتژی محور می‌توان از دیگر قابلیت‌های سنهاب نیز بهره‌برداری کرد. پایگاه داده سنهاب همچنین پتانسیلی برای کسب داده‌های از نوع تصویر و متن را نیز داراست که تحلیل چنین داده‌هایی دانش سودمندتری به‌دست می‌دهد. هریک از ماموریت‌های بیمه مرکزی را می‌توان براساس استراتژی داده‌محور و بر مبنای رویکرد SMART مورد بررسی قرار داد.

## منابع

۱. روحانی، سعید. حسینی، سمیه. تحلیل‌های عظیم داده، نقشه راه پیاده‌سازی، فناوری و ابزارها. انتشارات نیاز دانش. چاپ اول. ۹۴.
۲. قلاتی، یوسف. خاکور، مجید. فناوری اطلاعات موتور محرکه توسعه بازار بیمه. ۹۴.
3. BIG DATA WORKING GROUP, 2014, "Big Data Taxonomy", cloud security alliance.
4. Castro, D., Korte, T., 2013, "Data Innovation 101: An Introduction to the Technologies and Policies Supporting Data-Driven Innovation", Center for Data Innovation.
5. Patterson, T., 2015, "The Use of Information Technology in Risk Management", Complex Solutions Executive IBM Corporation.
6. Pipino, L.L., Lee, Y.W., & Wang, R.Y. , Apr 2002. Data quality assessment. Communications of the ACM, 45(4), 211–218. <http://dx.doi.org/10.1145/505248.506010>.
7. Han, J. and Kamber, M. , " Data Mining : Concepts and Techniques ", Second Edition , Morgan Kaufman Publisher, 2006.
8. The 2016 global data management benchmark report, 2016. Experian data quality.
9. Erl, th., Khattak, W., Buhler, P., "Big Data Fundamentals Concepts, Drivers & Techniques", 2015.