

ارائه یک مدل مفهومی جدید برای حوزه کلان‌داده‌ها و تحلیل کسب‌وکارهای داده‌محور در ایران بر مبنای مدل ارائه‌شده

مژگان فرهودی* رضوان کلانتر هرمزی** حسام زندحسامی***

* عضو هیات علمی، پژوهشگاه ارتباطات و فناوری اطلاعات، پژوهشکده فناوری اطلاعات

** عضو کارگروه ارزیابی و تشخیص صلاحیت شرکت‌ها و موسسات دانش بنیان

*** عضو هیات علمی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، دانشکده مدیریت و اقتصاد

تاریخ پذیرش: ۱۳۹۹/۱۱/۲۸

تاریخ دریافت: ۱۳۹۹/۰۴/۰۵

نوع مقاله: پژوهشی

چکیده

پیش‌بینی افزایش ده برابری داده‌های جهانی تا سال ۲۰۲۵ نشان می‌دهد که سفر داده‌ها برای سازمان‌ها تازه آغاز شده است. جمع‌آوری، ذخیره، تجزیه و تحلیل و استفاده از این طلای باارزش به نوآوری در شرکت‌ها و سازمان‌ها کمک می‌کند و مشاغل را به سمت آینده رقابتی سوق می‌دهد. هدف از این مقاله شناسایی و بررسی زمینه‌های مختلف ارزش‌آفرینی با محوریت کلان‌داده برای شرکت‌های نوپایی است که به داده‌ها به عنوان منبع اصلی کسب‌وکارشان تکیه دارند. در این راستا پس از بررسی ابعاد مختلف کلان‌داده، مدل مفهومی جدیدی در این حوزه ارائه می‌گردد. این مدل به شرکت‌ها کمک می‌کند تا بتوانند فعالیت‌های خود را با آگاهی و تمرکز بیشتری پیش ببرند. در ادامه، نتایج پیمایش شرکت‌های متقاضی دانش‌بنیان با کسب‌وکارهای داده‌محور ارائه می‌شود. نتایج بررسی‌ها نشان می‌دهد که اغلب شرکت‌ها بر روی دو حوزه «تحلیل» و «مصورسازی داده» تمرکز بیشتری نموده‌اند. همچنین، حوزه‌های «فین تک» و «رسانه» بیش از سایرین پذیرای فناوری‌ها بوده‌اند. از طرفی محصولات مربوط با «مدیریت و پردازش کلان‌داده‌ها» که شامل ارائه راهکار و راه‌اندازی خدمات پردازش و ذخیره‌سازی می‌شود، بیشترین حجم فروش را به خود اختصاص داده‌اند و میزان قابل توجهی از فعالیت شرکت‌ها بر روی محتوای اخبار، شبکه‌های اجتماعی و اطلاعات مربوط به معاملات بورس اختصاص یافته‌اند.

واژگان کلیدی: کلان‌داده، مدل مفهومی، کسب و کار داده‌محور، شرکت‌های دانش‌بنیان

۱ مقدمه

به باور بسیاری از صاحب‌نظران، علم داده جذاب‌ترین و پرکاربردترین علم قرن بیست و یکم است. در جهان امروز که فناوری حرف اول را می‌زند، همه افراد سؤال‌هایی دارند که تنها راه جواب دادن به آن‌ها بررسی دقیق و کامل حجم قابل‌توجهی از داده‌ها خواهد بود. از سازمان‌های تجاری کوچک‌تر گرفته تا مؤسسات دولتی، همه داده‌های نامحدودی در اختیار دارند که با دسته‌بندی، تفسیر و به‌کار گرفتن آن‌ها، می‌توان مقاصد گوناگونی را برآورده کرد؛ البته این کار ساده نیست و می‌تواند چالش‌برانگیز باشد. به عبارتی دیگر، سرمایه‌های اساسی در حال انتقال از نفت به داده است و به همین دلیل است که از داده به عنوان نفت جدید نام می‌برند. البته واضح است که داده نه تنها تمام شدنی نیست، بلکه با انجام هر تراکنش مالی، هر کلیک کاربر، نوشتن هر پیغام در شبکه‌های اجتماعی و بارگزاری هر تصویر جدید به حجم آن اضافه می‌شود. از سوی دیگر، تحلیل‌های نتیجه‌شده از تحلیل داده‌ها خود نیز داده‌های جدید هستند. حجم داده‌های موجود بصورت نمایی در حال افزایش است. همین حقیقت ساده نشان می‌دهد که ما تا چه حد به علم داده و دانشمند داده نیاز خواهیم داشت.

همانطور که گفته شد، در عصر حاضر، داده‌ها از ارزش بسیار زیادی برخوردارند و شرکت‌های مختلف با استخراج و تحلیل کلان‌داده‌ها به مزیت رقابتی دست می‌یابند؛ به عنوان مثال در خرده‌فروشی‌ها با تحلیل رفتار خرید مشتریان در آن واحد، می‌توان به سود بیشتری دست یافت. رشد داده که در سال ۲۰۲۰، ۴۰ زتابایت^۱ تخمین زده شده بود، در سال ۲۰۲۵ به ۱۷۵ زتابایت خواهد رسید که ۹۰ زتابایت آن توسط دستگاه‌های مربوط به اینترنت اشیا^۲ ایجاد می‌شود [۱]. همچنین پیش‌بینی می‌شود که تا سال ۲۰۲۵، هر فرد بطور متوسط روزانه ۵۰۰۰ تراکنش دیجیتال با دستگاه‌های مختلف خواهد داشت [۱] که این نشان‌دهنده حجم بسیار بالای تولید داده در آینده خواهد بود. از طرفی بنابر جمع‌بندی اتحادیه اروپا [۲]، اقتصاد داده با رشد ۱۲ درصدی از سال ۲۰۲۰ تا سال ۲۰۲۵ به ۱۰۵۴ میلیارد یورو خواهد رسید. تعداد شرکت‌ها و میزان اشتغال حوزه داده نیز که در سال ۲۰۱۸، به ترتیب ۲۰ و ۱۵ درصد از کل شرکت‌ها بوده است، سالانه با رشد چشمگیری مواجه هستند.

بهره‌بری از این موج جدید نیاز به سرمایه‌چندانی ندارد و بازار بسیار بکری برای فعالان اقتصادی فراهم آمده است. تنها مسئله حال حاضر، آشنا کردن صاحبان کسب‌وکارهای فناوری اطلاعات با سودهای سرشاری است که در شرکت‌هایشان راکد مانده است. پیشرفت تکنولوژی باعث کاهش هزینه جمع‌آوری اطلاعات و تحلیل آن توسط کسب و کارها شده است، لذا امروزه کسب‌وکارها می-

توانند با صرف هزینه و زمان کمتر نسبت به گذشته به اطلاعات دقیق‌تری دست یافته، تصمیم‌های آگاهانه‌تر و منطقی‌تری در مورد کسب و کار خود بگیرند. برخی از مزایایی که فناوری کلان داده‌ها برای کسب‌وکارها به همراه دارد عبارتند از [۳]:

- کاهش هزینه: فناوری‌های کلان‌داده مانند هادوپ^۳ و تجزیه و تحلیل مبتنی بر ابر می‌تواند کاهش چشمگیری در هزینه‌ها داشته باشد. به عنوان مثال، شرکت‌ها به جای پردازش و ذخیره مقادیر زیادی از داده‌های جدید در انبار داده‌ها، می‌توانند از امکانات هادوپ استفاده کرده و داده‌ها را در صورت نیاز به انبارهای سازمانی منتقل نمایند.

- تصمیم‌گیری سریع‌تر و بهتر: امروزه تمامی سازمان‌ها حتی سازمان‌های کوچک نیز دارای وب‌سایت و صفحه‌های اجتماعی هستند. اطلاعات مربوط به استفاده کاربران از این صفحات می‌تواند منجر به تشکیل بانک داده‌ای برای سازمان باشد. این بدان معنی است که تمامی سازمان‌ها می‌توانند با اجرای روش‌ها و الگوریتم‌های مناسب، اطلاعات مناسبی را در مورد فضای کسب‌وکار، مشتریان خود و سلاقی آنها به دست آورند و تصمیم‌گیری‌ها خود را بهتر و سریع‌تر انجام دهند.

- محصولات و خدمات جدید: کلان‌داده به شرکت‌ها توانایی ارائه محصولات و خدمات جدید با امکان دریافت بازخورد مشتری را داده است. سازمان‌ها برخلاف گذشته منتظر درخواست مشتری نیستند، بلکه از قبل نیاز مشتری را پیش‌بینی کرده و محصول را تولید می‌کنند. این مهم با کمک بررسی اطلاعات مربوط به عادات خرید مشتریان، سابقه جستجوهای آنها، و حتی سناریوهای فرضی به دست می‌آیند.

- توصیه محصول: با توجه به بررسی الگوهای موجود در جستجوهای اینترنتی فرد (و اطلاعات دیگری همچون محل‌هایی که فرد بیشتر به آنها مراجعه می‌کند)، می‌توان ذائقه‌ی وی در جستجوی‌های اینترنتی را شناسایی و پیش‌بینی کرده و محصولات یا محتوای جدید را به وی پیشنهاد داد، و بدین ترتیب از حضور وی در وب‌سایت مراقبت شود.

- تشخیص تقلب: سازمان‌هایی که فعالیت‌های مالی بزرگی دارند، همواره به دنبال راه‌هایی ارزان‌تر و موثرتر برای جلوگیری از سوءاستفاده‌های مالی هستند. برای مثال بیمه‌های درمانی سالانه هزینه‌های چشمگیری را به دلیل خسارت ناشی از کلاهبرداری‌های مختلف متحمل می‌شوند. در گذشته بیمه‌ها اطلاعات مربوط به صورت‌هزینه‌های مدعیان را با استانداردهای خود مقایسه کرده، به واقعی بودن یا نبودن ادعا پی می‌بردند، که این فرایند از چند هفته تا چند ماه زمان نیاز داشت. امروزه به کمک بررسی فاکتورهای مختلف، تشخیص ویدئویی، تشخیص مکالمه، و یادگیری ماشین (با

^۱ Zettabytes

^۲ Internet of Things (IOT)

^۳ Hadoop

بین مفاهیم می‌باشد. لازم به ذکر است برای تکمیل و پوشش حداکثری مفاهیم مرتبط در حوزه داده و تعیین ارتباطات بین آنها از نظرات خبرگان و صاحب‌نظران این حوزه نیز استفاده شده است. از طرفی برای بازنمایی دانش ذکر شده از مدل درختواره استفاده شده است. لذا هدف از این مقاله ارائه یک تصویر کلان از ابعاد مختلف موضوع کلان داده است که از مدل مذکور برای ترسیم مفاهیم و ارتباطات بین آنها استفاده شده است.

بطور کلی، یکی از مدل‌های بازنمایی دانش، مدل بازنمایی مبتنی بر هستان‌شناسی (یا درختواره) است [۱۱]. فعالیت‌های زیادی در خصوص ارائه مدل مفهومی از یک مسئله یا موضوع انجام شده است ولی تعداد محدودی از آنها جهت بازنمایی دانش حوزه از مدل مفهومی مبتنی بر هستان‌شناسی استفاده کرده‌اند [۱۱] تا [۱۴]. [۱۵] و [۱۶]. در تمامی موارد یاد شده، مدل مفهومی موضوع مربوطه ترسیم شده و در برخی از آنها با استفاده از مدل مفهومی به استخراج پروژها و یا حوزه‌های فعالیت ارائه شده است. در [۱۷] با ارائه یک مدل مفهومی متشکل از دو سطح فنی و غیرفنی به معرفی جامعی از روندها و حوزه‌های تحقیقاتی نوین داده‌های عظیم (کلان داده‌ها) پرداخته شده است.

۳ مدل مفهومی حوزه داده

در این بخش به تشریح مدل مفهومی پیشنهادی در حوزه کلان داده پرداخته و ابعاد مختلف آن را از منظرهای مختلف بررسی خواهیم کرد. مدل مفهومی در واقع چارچوبی است که در آن به شکلی نظام‌مند نیازمندی‌های فنی و محیطی یک موضوع بیان می‌شود [۱۱]. برای بازنمایی این مدل مفهومی لازم است تا شناخت خوبی بر روی نیازمندی‌های موضوع داده وجود داشته باشد تا بتوان این نیازمندی‌ها را با فرایندها، روش‌ها، ابزار و استانداردهای موجود تطبیق داد. مدل مفهومی یک واژه شناخته شده است و عموماً به درک مفهومی مسئله‌ای قبل از تحلیل و طراحی اطلاق می‌گردد. در این بخش ابتدا ابعاد مختلف حوزه داده معرفی می‌شوند و پس از آن مدل مفهومی مبتنی بر درختواره ارائه خواهد شد.

۱.۳ ابعاد مختلف مطرح در حوزه کلان داده

همانطور که بطور ضمنی اشاره شد، موضوع داده و توسعه کاربردهای آن، ابعاد مختلفی دارد. در شکل ۱، جنبه‌های گوناگون این مسئله نمایش داده شده است که در ادامه هر یک از این ابعاد و مباحث پیرامون آن مورد بررسی قرار گرفته است:

کمک کلان داده) به شناخت الگوهای تقلب پرداخته و از ضررهای ناشی از سوءاستفاده‌های مالی می‌کاهند. در این مقاله، ابتدا به مزایای استفاده از کلان داده‌ها در کسب و کارها پرداخته و سپس مدل مفهومی حوزه داده ارائه می‌گردد. در انتها نیز یک پیمایش جامع بر روی شرکت‌های متقاضی دانش‌بنیانی در داخل کشور که محوریت کسب و کار آنها بر روی داده می‌باشد انجام شده و نتایج بررسی ارائه می‌گردد.

۲ کارهای انجام شده

طبق بررسی‌های بعمل آمده، پژوهش‌های متعددی به ارائه ابعاد داده از منظرهای مختلف پرداخته‌اند. به عنوان مثال در [۴] یک درختواره^۱ برای مدل‌های کسب و کار حوزه داده ارائه شده که از اطلاعات شرکت‌های استارت‌آپ برای تهیه این مدل استفاده نموده است. [۵] با تمرکز بر چالش‌های حوزه داده به ارائه مدل مفهومی این حوزه پرداخته است و تلاش کرده تا سه بعد مرتبط بودن، جامعیت و اولویت‌های نسبی را در آن نشان دهد. هدف از ارائه مدل مربوطه، افزایش کیفیت فرایند تصمیم‌گیری می‌باشد. در مقاله [۶]، هدف اصلی توسعه یک مدل جامع برای کلان داده می‌باشد که شامل عواملی است که در موفقیت یا عدم موفقیت اجرای کلان-داده‌ها در سازمان‌ها نقش دارند. و لذا مدل پیشنهادی در این مقاله عوامل مهمی را در اختیار مدیران فناوری اطلاعات و تصمیم‌گیران قرار می‌دهد که برای اطمینان از دستیابی به مزیت رقابتی باید در هنگام تصمیم‌گیری برای استفاده از کلان داده در نظر بگیرند. همچنین در [۷] و [۸] و [۹]، درختواره یا عبارتی ابعاد مختلفی برای کلان داده براساس مقالات مرتبط ارائه داده‌اند. در مقاله [۱۰] نیز با هدف تاکید بر ذخیره‌سازی، پیش‌پردازش، پردازش و امنیت، به ارائه یک درختواره برای مدیریت حوزه کلان داده پرداخته است. نتایج پژوهش‌های انجام شده نشان می‌دهد که هر یک از مدل‌های ارائه شده در این پژوهش‌ها، تمرکز خود را بر روی بخشی از موضوع کلان داده گذاشته‌اند؛ بعنوان مثال مدل ارائه شده در [۵] با تمرکز بر تصمیم‌گیری در سازمان و افزایش کارایی آن به ایجاد و ترسیم مدل پرداخته است و لذا تمام جنبه‌های حوزه کلان داده را شامل نمی‌شود. اما در مقاله حاضر، سعی شده است با دید کلی‌تری به موضوع نگاه شود و به منظور درک دانش و مفاهیم مطرح در این حوزه به ترسیم یک تصویر کلان^۲ و یا نمایش مفاهیم این حوزه و ارتباطات بین آنها با عنوان مدل مفهومی پرداخته شود. روش مورد استفاده در این مقاله براساس مطالعه پژوهش‌های مرتبط با حوزه کلان داده و استخراج مفاهیم این حوزه و سپس تعیین ارتباطات

^۱ Taxonomy

^۲ Big Picture

۱،۲،۳ انواع داده

داده‌های موجود در فضای مجازی اعم از وب و شبکه‌های اجتماعی، به شکل‌های مختلفی تولید شده و در دسترس وجود دارد. این انواع عبارتند از داده‌های متنی، صوتی (یا گفتاری)، تصویری و ویدئویی.

۲،۲،۳ قالب داده

همانطور که در شکل ۲ نیز مشخص است، قالب داده‌ها را می‌توان در سه دسته ساخت‌یافته، نیمه‌ساخت‌یافته و غیرساخت‌یافته یا بدون ساختار دسته‌بندی کرد [۱۸]. هر داده‌ای که قابلیت ذخیره شدن، دسترسی و پردازش را داشته باشد و به یک قالب ثابت باشد، به عنوان یک «داده ساخت‌یافته» در نظر گرفته می‌شود. مقابل این تعریف، داده‌های غیرساختاری می‌شوند. داده‌های نیمه ساخت‌یافته می‌توانند هر دو نوع داده (ساخت‌یافته و غیرساخت‌یافته) را شامل شوند. ما می‌توانیم داده‌های نیمه ساخت‌یافته را به صورت ساخت-یافته ببینیم اما در صورتی که بخواهیم آنها را به عنوان یک جدول رابطه‌ای در نظر بگیریم، در DBMS تعریف نشده است. به عنوان مثالی از آن می‌توان به داده‌های موجود در یک فایل اکسل اشاره نمود.

۳،۲،۳ منابع داده

منابع داده در کسب‌وکارهای داده‌محور را می‌توان به دو دسته منابع داده داخلی و خارجی تقسیم کرد. داده‌های داخل شرکت شامل داده‌هایی است که یا در شرکت موجود هستند و یا توسط شرکت تولید می‌شوند. تولید داده در داخل شرکت یا از طریق کارمندان و بوسیله سنسورها و ابزار ردیابی صورت می‌گیرد و یا از طریق جمع‌سپاری^۱ بدست می‌آید. داده‌های خارجی نیز به سه دسته داده‌های ارائه‌شده توسط مشتری، داده‌های رایگان و داده‌های اکتساب شده از عامل سوم تقسیم می‌شود. داده‌های رایگان خود شامل داده‌های باز در دسترس، داده‌های بدست آمده از شبکه‌های اجتماعی و داده‌هایی که از خزش وب بدست می‌آیند می‌باشد [۴].

۴،۲،۳ کاربرد داده

امروزه داده به سرعت در حال گسترش است و بدلیل استفاده روزافزون از سامانه‌های دیجیتال و منابع تولید اطلاعات، کاربرد داده در تمامی حوزه‌هایی که با داده‌ها و تحلیل آنها سر و کار دارند، گسترش یافته است و نتایج بسیار قابل توجهی نیز منتج شده است [۷]. با توجه به تعدد حوزه‌های کاربردی داده‌ها، ذکر تمامی کاربردهای موجود امکان‌پذیر نیست و در شکل ۲ به برخی از این موارد از قبیل حوزه‌های بانکداری، سلامت، حمل و نقل، آموزش، بورس، رسانه و سرگرمی، تبلیغات و غیره اشاره شده است [۱۹].



شکل ۱. ابعاد مختلف حوزه کلان داده

- انواع: در این بعد، به انواع داده‌های تولید شده در فضای مجازی، سازمان‌ها و برنامه‌های مختلف پرداخته می‌شود.
- قالب: در این بخش به قالب داده‌های گردآوری و یا تولید شده توسط سازمان‌ها، افراد، ابزارها و یا نرم‌افزارها اشاره می‌شود.
- کاربرد: داده‌ها معمولاً با هدف کاربردهای مختلف مورد تحلیل یا پردازش قرار می‌گیرند که در این بخش، به این کاربردها اشاره شده است.
- فناوری: امروزه بدلیل اهمیت داده، بسیاری از فناوری‌های روز دنیا مبتنی بر داده هستند که مصادیق آن در این بخش نام برده شده است.
- ذی‌نفعان: در این بعد به ذی‌نفعان این حوزه از جمله مشتریان و نقش‌آفرینان که در واقع همان تولیدکنندگان و صاحبان داده، کاربران داده و نیز توسعه‌دهندگان خدمات هستند اشاره می‌گردد.

۲،۳ مدل مفهومی مبتنی بر درختواره

جهت تهیه مدل مفهومی این حوزه تلاش گردید تا با مطالعه تطبیقی نمونه‌های موجود و بررسی پژوهش‌های انجام شده و همچنین مشورت با افراد متخصص و صاحب‌نظر این حوزه، به بررسی ابعاد مختلف ذکر شده در بالا بپردازیم. شکل ۲، مدل مفهومی تهیه شده حوزه داده را نشان می‌دهد. در این شکل به منظور بازنمایی دانش این حوزه، از درختواره استفاده شده است. همان‌گونه که پیش از این بیان شد برای تعیین روابط اصلی و گره-های مربوطه آن در ترسیم مدل مفهومی ابتدا ابعاد مختلف این حوزه شناسایی شد و در مرحله بعد، هر یک از این جنبه‌ها بسط داده شدند. برای آشنایی بیشتر با مباحث مطروحه در هر یک از ابعاد موضوع داده در ادامه به شرح هر یک از آنها پرداخته می‌شود.

^۱ Crowdsourcing

۵.۲.۳ فناوری‌های مرتبط

با توجه به فرصت‌های کسب و کاری حوزه و همچنین اهمیت روزافزون داده‌های تولید شده از ابزارها و یا افراد مختلف، می‌توان گفت داده، محوریت اغلب فناوری‌های روز دنیا را تشکیل می‌دهد. به عنوان مثال فناوری‌هایی از قبیل اینترنت اشیا، بیومتریک، زنجیره بلوکی، واقعیت مجازی، هوش مصنوعی و ... در زمره این فناوری‌ها قرار داد.

۶.۲.۳ ذینفعان

ذینفعان این حوزه را به طور کلی می‌توان در دو گروه مشتریان و نقش‌آفرینان دسته‌بندی نمود [۱۲] و [۱۵]. در ادامه تعریف هر یک از این دو گروه و نقش‌های مؤثر آنها فهرست شده است. لازم به توضیح است که بسیاری از شرکت‌ها می‌توانند نقش‌های مختلفی را ایفا نمایند به عنوان مثال یک شرکت توسعه‌دهنده خدمات می‌تواند خود تولیدکننده داده و از طرفی مشتری یکی از خدمات حوزه داده‌ای باشد.

الف) مشتریان: مشتریان حوزه داده را می‌توان در دو گروه حقیقی و حقوقی دسته‌بندی کرد:

- حقیقی: این گروه از مشتریان، در واقع کاربران نهایی یا همان استفاده‌کنندگان داده یا خدمات مبتنی بر داده هستند. افرادی از قبیل گردشگران، مدیران، کاربران عمومی که اغلب از نتایج تحلیل و پردازش داده‌ها استفاده می‌کنند یا پژوهشگرانی که علاوه بر استفاده از داده‌های پردازش شده، خود می‌توانند با استفاده از داده‌های خام، به تحلیل آنها پرداخته و اطلاعات و یا دانش مربوطه را ایجاد نمایند.
- حقوقی: کاربران حقوقی شامل سازمان‌ها و شرکت‌هایی هستند که به عنوان مصرف‌کننده داده چه به لحاظ استفاده از خود داده و چه به لحاظ استفاده داده جهت پردازش و تحلیل محسوب می‌شوند. شرکت‌های خصوصی و بین‌المللی، بانک‌ها، مراکز پژوهشی مصادیقی از این نوع مشتریان هستند.

ب) نقش‌آفرینان: معمولاً به سه دسته حاکمیت، صاحبان و یا تولیدکنندگان داده و همچنین ارائه‌دهندگان خدمات مبتنی بر داده تقسیم می‌شوند.

شایان ذکر است موارد ذکر شده در شکل ۲ تنها مصادیقی از انواع مشتریان و نقش‌آفرینان هستند که مسلماً می‌توان برای هر دسته نمونه‌های بیشتری را نیز نام برد.

۷.۲.۳ مدل درآمد

هر شرکت برای دوام و ماندگاری در طولانی مدت، حداقل نیاز به یک مدل درآمدی دارد. مدل‌های درآمدی در کسب‌وکارهای داده-محور عبارتند از: فروش دارایی، وام دادن و اجاره دادن، صدور

مجوز، هزینه بر حسب استفاده، حق عضویت، حق کمیسیون و تبلیغات [۴].

۸.۲.۳ حوزه‌های فعالیت داده

در مدل مفهومی ارائه شده، باید گفت بخشی که استارت‌آپ‌های حوزه داده می‌توانند از داده خلق ارزش یا ثروت نمایند، همان بخش مربوط به حوزه‌های فعالیت داده به همراه زیرمجموعه‌های مطرح شده در آنها می‌باشد که عبارتند از: تولید و جمع‌آوری داده، ذخیره و بازیابی، تحلیل، مصورسازی، حکمرانی داده و بازکردن داده.

الف) تولید/جمع‌آوری داده

معمولاً داده‌ها بطور مستمر در فضای مجازی از طریق افراد و یا ابزارهای مختلف در حال تولید و انتشار هستند؛ لذا جمع‌آوری داده‌های پراکنده موجود در سایت‌ها و یا سازمان‌های مختلف و دسته‌بندی آنها به منظور انجام پردازش از فعالیت‌های مهمی است که در حوزه داده انجام می‌شود. برخی شرکت‌های مطرح دنیا تنها به پردازش بر روی آرشیو اطلاعات و مستندات موجود در وب پرداخته و سپس آن را بر اساس نیاز در اختیار شرکت‌های تحلیلگر قرار می‌دهند. منابع تولید داده‌ها بسیار مختلف و متنوع است. داده معمولاً از روش‌های مختلف تولید و یا جمع‌آوری می‌شود که عبارتند از: حسگرهای اینترنت اشیا، بستر جمع‌سپاری، خزش وب و غیره.

ب) ذخیره و بازیابی داده

ذخیره و بازیابی کلان‌داده‌ها به دلیل ویژگی‌های کلان‌داده از قبیل حجم، سرعت تولید، تنوع، صحت و غیره از اهمیت زیادی برخوردار است و چالش‌های زیادی را به دنبال دارد. به عنوان مثال، داده‌ها در کلان‌داده بصورت جریانی تولید می‌شود. داده جریانی به داده‌ای گفته می‌شود که دارای فرایند پردازشی بصورت بلادرنگ است. موضوع ذخیره‌سازی در تمامی فرایندهای از قبیل استخراج داده، پیش‌پردازش، پردازش و انتقال داده نقش بازی می‌کند.

ذخیره‌سازی داده‌ها می‌تواند بصورت رابطه‌ای (مانند SQL) و یا غیررابطه‌ای (مانند No SQL) باشد که در مبحث کلان‌داده عمدتاً از روش غیررابطه‌ای برای ذخیره استفاده می‌شود. روش غیررابطه‌ای خود شامل انواع مختلفی از قبیل ذخیره‌سازی مبتنی بر سند، مبتنی بر ستون، مبتنی بر گراف و مبتنی بر کلید/مقدار می‌باشد [۸].

ج) تحلیل داده

با گسترش روزافزون استفاده از فناوری اطلاعات و نرخ بالای تولید اطلاعات دیجیتال، توسعه فناوری‌ها و روش‌های تحلیل کلان‌داده‌ها از اهمیت بسزایی برخوردار شده‌اند. با توجه به میزان خودکار بودن

هدف از حکمرانی یا حاکمیت داده این است که به سازمان کمک کند تا داده‌های خود را به عنوان یک دارایی مهم سازمانی مدیریت نماید.

حکمرانی داده، مجموعه‌ای از اصول، سیاست‌ها، فرایندها، چارچوب‌ها و سنجه‌ها را فراهم می‌آورد و سعی می‌کند تا بر مدیریت داده‌ها نظارت کند و فعالیت‌های مدیریت داده را در تمامی سطوح هدایت کند. حکمرانی داده شامل بخش‌های متنوعی است [۲۲] که در شکل ۱۱ نشان داده شده است. برخی از این موارد عبارتند از: یکپارچه‌سازی داده، معماری داده، مدیریت محتوا و اسناد، مدلسازی و طراحی داده، ذخیره‌سازی و عملیات داده و ...

ی) توزیع و باز کردن داده

تا چند سال پیش، اشتراک داده‌ها و یا باز کردن آنها ایده‌ای غیرقابل تصور بود؛ زیرا صاحبان داده نگران از بین رفتن زحمات‌ها و یا مزیت رقابتی خود بودند. اما امروزه با تمرکز جامعه علمی بر روی دسترسی آزاد، محققان متوجه مزایا و نیاز به دسترسی و اشتراک-گذاری آزاد داده‌ها شده‌اند. با گذشت زمان، موسسات، انتشارات، حامیان مالی و محققان بیشتری به مزایای دسترسی آزاد اطلاعات پی می‌برند. نمونه‌ای از انواع داده‌هایی که می‌تواند بصورت آزاد در دسترس قرار گیرد عبارت است از: داده‌های علمی و پژوهشی دانشگاه‌ها و مراکز پژوهشی، داده‌های سازمانی و داده‌های دولتی.

از طرفی بسته به نیاز متقاضیان داده، هر یک از انواع داده‌ها را می‌توان به یکی از صور خام (بدون هیچگونه پردازشی بر روی داده)، داده‌های پردازش شده و یا داده‌های گرافیکی (مانند اینفوگرافی یا نمودارهای مختلف) و یا به هر سه شکل آذارسازی نمود.

۴ بررسی و تحلیل کسب‌وکارهای داده‌محور در کشور

همانطور که در مدل مفهومی پیشنهادی نشان داده شد، حوزه‌های فعالیت بر روی داده‌ها عبارتند از: تولید و جمع‌آوری داده، ذخیره و بازیابی، تحلیل، مصورسازی، حکمرانی و توزیع و باز کردن داده.

فرایند تجزیه و تحلیل می‌توان آن را به چهار گروه توصیفی، تشخیصی، پیشگویانه و تجویزی تقسیم نمود [۲۰]. در تجزیه و تحلیل توصیفی، هدف پاسخ به این نوع سوالات است: چه اتفاقی افتاده است و یا چه چیزی دارد اتفاق می‌افتد. در این روش با دستیابی به سوابق که ریشه در گذشته و حال دارد، امکان تحلیل داده مبتنی بر رویدادهای گذشته و حال و همچنین تولید بینش میسر می‌شود. از این نوع تحلیل می‌توان به خوشه‌بندی اشاره نمود. در تجزیه و تحلیل تشخیصی، هدف اصلی پاسخ به این نوع سوالات است: چرا این اتفاق افتاده است و یا چرا دارد این اتفاق می‌افتد؟ روند به چه صورت است؟ چه الگویی وجود دارد؟ در این روش با بررسی سوابق داده، بینش ایجاد می‌گردد. تجزیه و تحلیل احساسات، تحلیل لاگ، کشف الگوهای تقلب و غیره مصادیقی از این نوع تحلیل هستند.

در تجزیه و تحلیل پیشگویانه، هدف اصلی پاسخ به این نوع سوالات است: چه چیزی اتفاق خواهد افتاد؟ چگونه می‌توان واکنش نشان داد؟ در این نوع، با تحلیل داده و تولید بینش امکان پیش‌بینی داده جدید در رابطه با حال و آینده وجود دارد. پیش‌بینی رفتار مشتری نمونه‌ای از این نوع تحلیل است. در تجزیه و تحلیل تجویزی، هدف اصلی پاسخ به این نوع سوالات است: چه کاری بایستی انجام شود تا این اتفاق بیفتد.

از طرفی، روش‌های متعددی برای تحلیل وجود دارد [۹] که به عنوان نمونه می‌توان یادگیری ماشین و پردازش زبان طبیعی را برشمرد. روش‌های یادگیری ماشین بصورت خودکار و مقیاس پذیر اجازه بینش داده‌های کلان و چندبعدی را امکان‌پذیر می‌سازند.

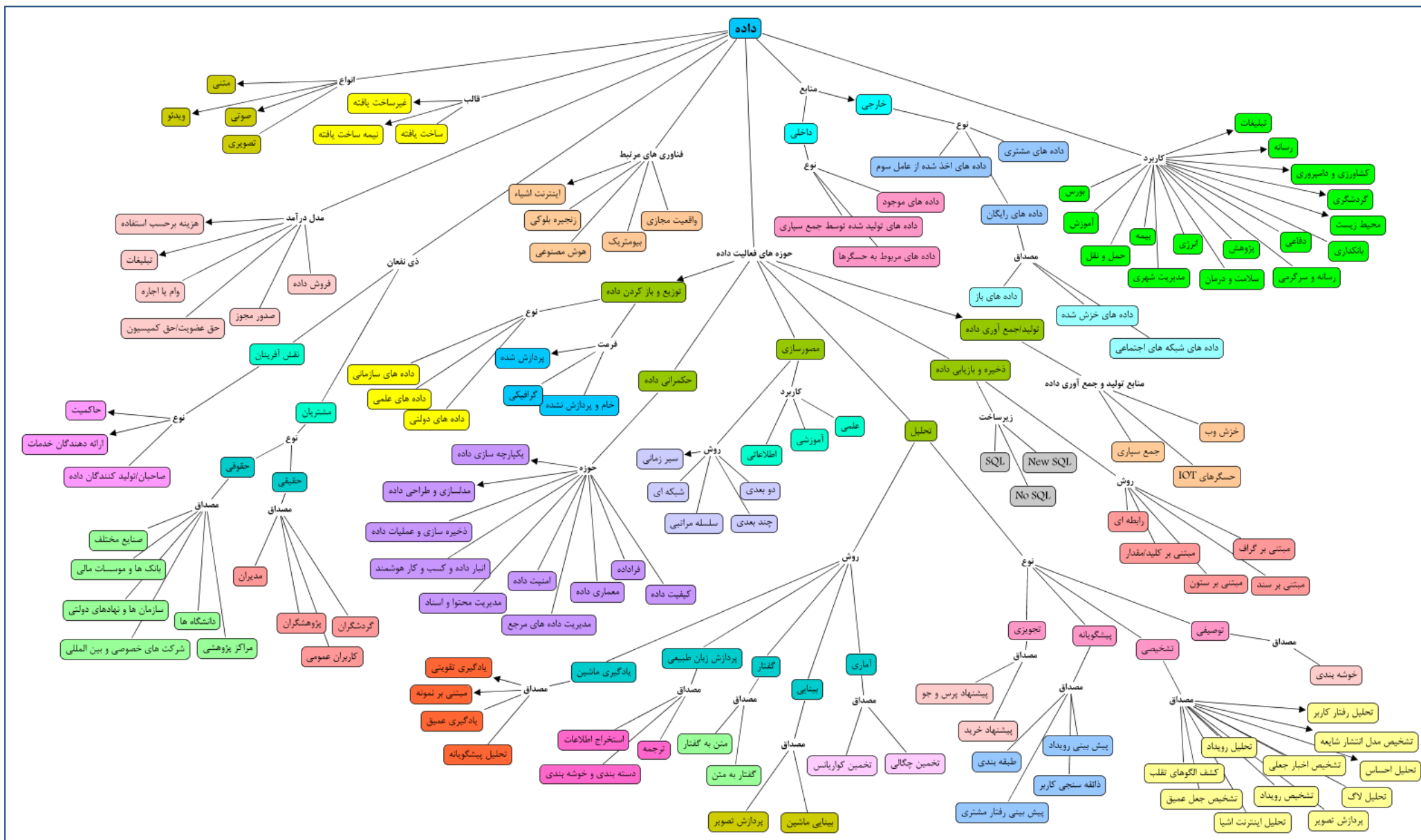
د) مصورسازی داده

مصورسازی در واقع عبارت است از شیوه‌های بازنمایی و ارائه داده یا اطلاعات در قالب‌های تصویری. این روش نمایش با بهره‌گیری از فنون خاص انجام می‌پذیرد و قادر است اطلاعات را در یک فرایند دیداری با قدرت بیشتر در درک و شناخت آن عرضه نماید.

مصورسازی عمدتاً در جهت کاربردپذیر ساختن داده‌ها است [۲۱]. کاربردهای مختلفی برای این امر وجود دارد. نکته‌ای که در رابطه با تفاوت میان مصورسازی علمی و مصورسازی اطلاعات باید بیان کرد این است که مصورسازی علمی عمدتاً به منظور کاربردپذیر ساختن داده‌های فیزیکی بکار می‌رود، در صورتیکه مصورسازی اطلاعات، به شیوه‌های تصویرسازی عملی داده‌های انتزاعی می‌پردازد.

همچنین مصورسازی می‌تواند به یکی از روش‌ها یا شکل‌های دوبعدی، چندبعدی، سلسله مراتبی، شبکه‌ای و سیر زمانی صورت پذیرد.

ه) حکمرانی داده



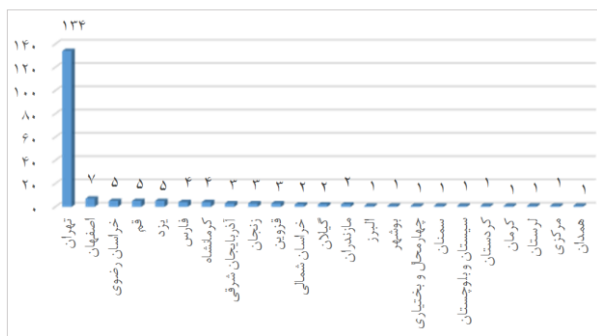
شکل ۲. مدل مفهومی حوزه داده مبتنی بر درختواره

در این مقاله، برای بررسی وضعیت کسب و کارهای داده‌محور در کشور، یک مطالعه موردی بر روی شرکت‌های متقاضی دانش‌بنیانی این حوزه انجام شد. در این خصوص، ۲۷۵ محصول فعال در کشور که توسط ۱۸۹ شرکت توسعه داده شده‌اند، مورد بررسی قرار گرفته است.

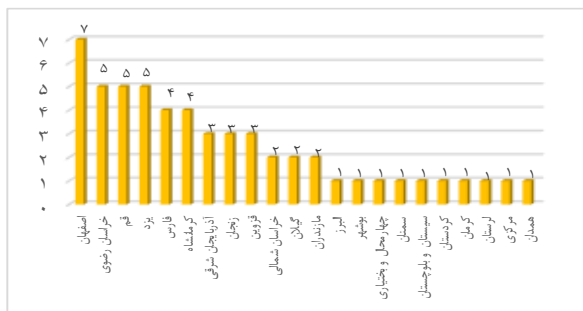
مبنای استخراج محصولات، دسته‌بندی جامع و استاندارد تصویب شده در «کارگروه ارزیابی و تشخیص صلاحیت شرکت‌ها و موسسات دانش بنیان حوزه فناوری اطلاعات» بوده است که از طریق سامانه دانش‌بنیان^۱ در دسترس عموم است.

بر اساس این سطح‌بندی، دسته‌هایی که محصولات و خدمات مرتبط با پردازش و تحلیل داده‌ها را تحت پوشش دارند، انتخاب شده‌اند. تحلیل آماری در دو سطح ۳ و ۴ دسته‌بندی محصولات دانش بنیان حوزه فناوری اطلاعات انجام شده است. نتایج این بررسی همراه با جزئیات در ادامه ارائه شده است.

شکل ۳، سیر زمانی ثبت شرکت‌های فعال در حوزه داده را در کشور نشان می‌دهد. همانطور که در شکل مشخص است، از سال ۹۳ الی ۹۶، بیشترین نرخ ثبت شرکت‌ها در این حوزه را داشته‌ایم. شاید بتوان دلیل عدم رشد ثبت این شرکتها در سالهای اخیر را نرخ بالای تورم و هزینه‌بر بودن خرید سخت‌افزار و زیرساخت‌های ارائه سرویس، تحریم‌ها و ... دانست که متأسفانه مانعی برای ورود افراد توانمند به این حوزه شده است.

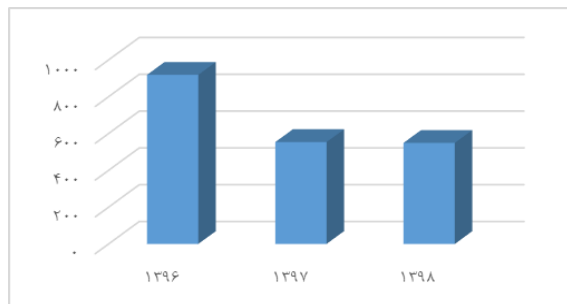


شکل ۴. پراکندگی استانی شرکت‌ها



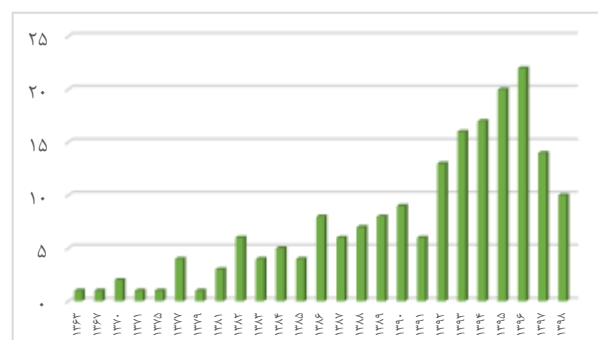
شکل ۵. پراکندگی استانی شرکت‌ها (بدون در نظر گرفتن تهران)

در شکل ۶ مجموع درآمد شرکت‌های دارای محصولات حوزه داده به تفکیک سال مشخص شده است. این درآمد بر اساس اظهارنامه ارائه شده شرکت‌ها محاسبه شده است.



شکل ۶. درآمد اظهار شده شرکت‌ها به تفکیک سال (میلیارد تومان)

همانطور که در شکل فوق هم مشخص است، درآمد اظهار شده شرکت‌ها در سال ۱۳۹۶ بیش از سال‌های ۱۳۹۷ و ۱۳۹۸ بوده است. متأسفانه آمار نشان می‌دهد که در یکی دو سال اخیر بدلیل اعمال تحریم‌های بیشتر از جمله تحریم‌های سخت‌افزاری و نرم-افزاری و نیز مشکلات اقتصادی شرکت‌ها، فعالیت آنها چه به لحاظ رونق تولید محصولات و چه به لحاظ درآمدزایی با افت محسوسی مواجه شده است.



شکل ۳. سیر زمانی ثبت شرکت‌های حوزه داده

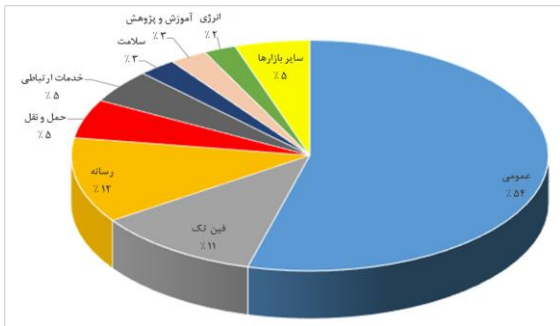
همچنین این شرکت‌ها در ۲۳ استان کشور، ثبت و در حال فعالیت هستند (از سایر استان‌ها، اطلاعاتی در این خصوص ثبت نشده است). پراکندگی جغرافیایی این شرکت‌ها در شکل ۴ نشان داده شده است.

بدلیل تعداد بالای شرکت‌ها در استان تهران، در شکل ۴ وضعیت استان‌های دیگر خیلی واضح نیست؛ لذا صرف‌نظر از استان تهران که به تنهایی دارای ۱۳۴ شرکت در حوزه داده است، وضعیت مابقی استان‌ها در شکل ۵ ارائه شده است.

^۱ <http://daneshbonyan.isti.ir/>

نکته جالب توجه در دو نمودار ارائه شده در شکل ۹، میزان فروش محصولات شرکت‌های زیر ۳ سال و بین ۳ تا ۱۰ سال است. نتایج نشان می‌دهد علیرغم نزدیک بودن تعداد محصولات در این دو گروه، فروش حاصل شده در شرکت‌های بین ۳ تا ۱۰ سال، حدود صد برابر فروش محصولات در شرکت‌های نوپا است.

بررسی دیگری که بر روی اطلاعات مستخرج صورت گرفت، بازار هدف محصولات است؛ بدین مفهوم که آیا این محصولات برای کاربرد عام توسعه داده شده‌اند و یا بازار خاصی برای فروش و ارائه خدمت مدنظر ارائه‌کنندگان آنها بوده است. بر این اساس و با مطالعه ویژگی‌های تک‌تک محصولات مشخص شد که در بین بازارهای عمودی، دو حوزه «فین‌تک» و «رسانه» بیش از سایرین مورد توجه بوده است.



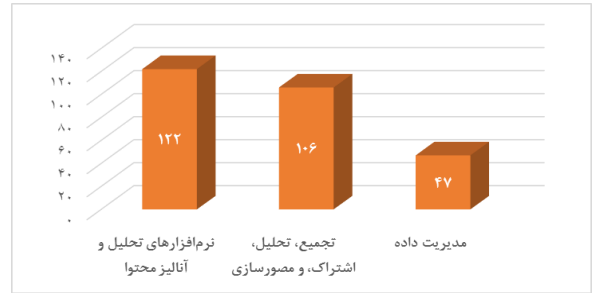
شکل ۱۰: سهم هریک از حوزه‌های کاربردی محصولات

شکل ۱۱ نیز نشان می‌دهد که در دسته «نرم‌افزارهای تجمیع، تحلیل، اشتراک و مصورسازی» درصد بیشتری از محصولات برای استفاده در بازارهای خاص توسعه داده شده‌اند.



شکل ۱۱: تفکیک حوزه‌های کاربردی محصولات توسعه داده شده

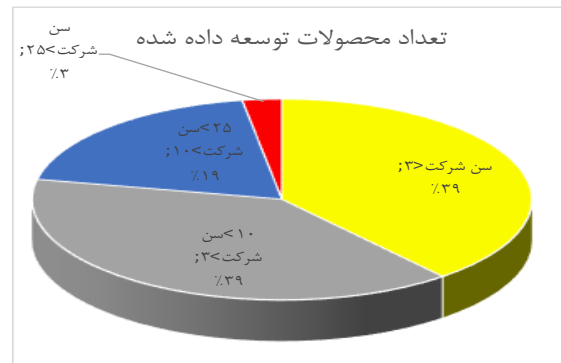
همانطور که در ابتدای این بخش توضیح داده شد، هر یک از سه دسته مورد بررسی در بالا، خود شامل زیردسته‌هایی است که ویژگی‌ها و کارکرد محصولات را بیشتر نمایان می‌سازد. لذا در ادامه، آمار مربوط به هریک از زیردسته‌ها بصورت مجزا ارائه شده است.



شکل ۷: تعداد محصولات حوزه داده به تفکیک دسته‌های مختلف

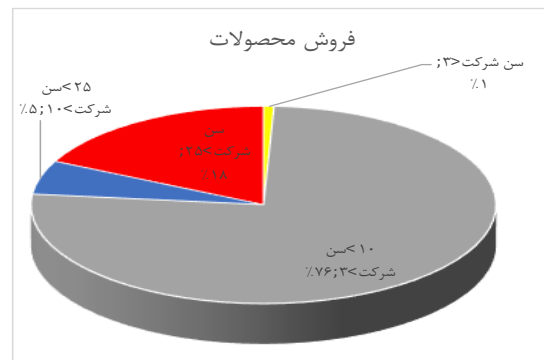
شکل ۷، تعداد محصولات هر یک از سه دسته محصولات مرتبط با داده را نمایش می‌دهد. همانطور که در نمودار نمایان است، بدلیل اهمیت و کاربرد روزافزون مباحث مربوط به تحلیل و مصورسازی داده‌ها، نرم‌افزارهای توسعه داده شده در این دو دسته نیز تعداد قابل توجهی را به خود اختصاص داده‌اند که می‌تواند بیان‌کننده بلوغ نسبی محصولات موجود در کشور در هر دسته باشد.

شکل ۸، بیانگر سن شرکت‌های توسعه‌دهنده محصولات است. همانطور که در شکل مشخص است، درصد شرکت‌های با رده سنی زیر ۱۰ سال بیش از سایرین است که میزان توجه به محصولات این حوزه با زمان ظهور و گسترش فناوری‌های پردازش و تحلیل داده مطابقت دارد.



شکل ۸: تعداد محصولات شرکت‌ها در بازه‌های سنی مختلف

در شکل ۹ مقایسه‌ای میان تعداد و فروش محصولات با توجه به سن شرکت‌های توسعه‌دهنده محصولات، انجام شده است.



شکل ۹: میزان فروش محصولات با توجه به سن شرکت‌ها

ارائه یک مدل مفهومی جدید برای حوزه کلان داده‌ها و تحلیل کسب و کارهای داده‌محور در ایران بر مبنای مدل ارائه شده

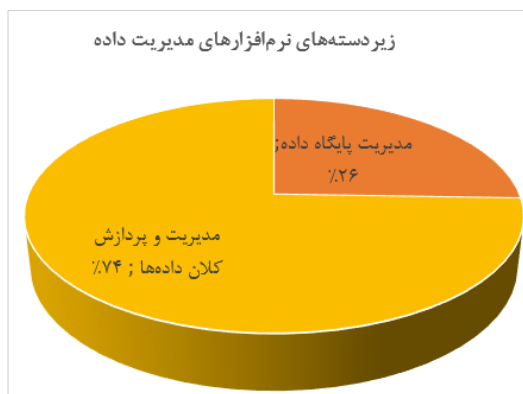
تجمع داده		
ابزارها و خدمات داده‌کاوی	۴۸	۵,۸۳٪
هوش تجاری	۵۷	۲۳,۹۶٪
مدیریت پایگاه داده	۱۲	۶,۰۲٪
مدیریت و پردازش کلان داده‌ها	۳۵	۵۵,۰۹٪
نرم‌افزارهای آنالیز متن	۶۳	۷,۰۴٪
نرم‌افزارهای آنالیز و سنتز صدا	۲۴	۰,۲۳٪
نرم‌افزارهای آنالیز ویدئو و تصویر	۲۵	۱,۰۶٪
ابزار جستجو و توصیه‌گر	۱۰	۰,۴۵٪

جدول ۱ نشان می‌دهد که زیردسته «مدیریت و پردازش کلان داده‌ها» که شامل ارائه راهکار و راه‌اندازی خدمات پردازش و ذخیره سازی کلان داده‌ها می‌شود، بیشترین حجم فروش را به خود اختصاص داده است. راه‌اندازی و ارائه خدماتی مانند Spark, Hadoop و Zookeeper در این دسته قرار دارند.

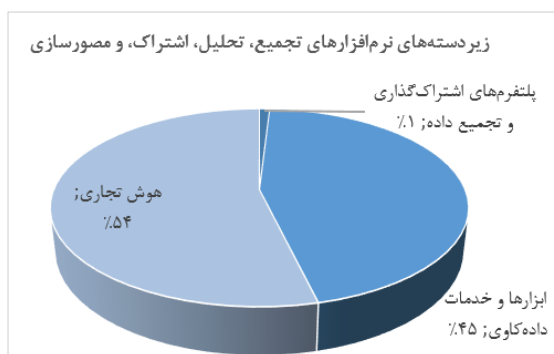
نکته لازم به ذکر در رابطه با حوزه کاربردی محصولات، استفاده قابل توجه از محصولات دو زیردسته «ابزارها و خدمات داده‌کاوی» و «نرم‌افزارهای آنالیز متن» به ترتیب در حوزه‌های فین تک و محصولات رسانه‌ای است. طبق بررسی‌های صورت گرفته، از بین محصولات ارزیابی شده در این دو زیردسته، ۳۴ درصد از نرم افزارهای تحلیل متن در حوزه رسانه با محوریت خزش اخبار و شبکه‌های اجتماعی و ۳۵ درصد از ابزارها و خدمات داده‌کاوی در حوزه فین تک با محوریت الگوریتم‌های معاملاتی بورس و سرمایه گذاری توسعه داده شده‌اند.

۵ نتیجه‌گیری و کارهای آتی

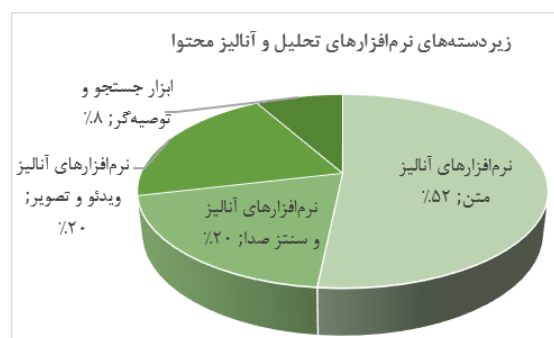
امروزه بسیاری از شرکت‌ها با تحلیل داده‌ها، ایجاد پایگاه‌های داده، مدیریت دانش و فروش خدمات مبتنی بر داده به درآمدهای هنگفتی رسیده‌اند. همانطور که می‌دانیم روزانه تصمیمات زیادی در سطح خرد و کلان گرفته می‌شود که این تصمیم‌ها اثرات زیادی بر موفقیت کسب و کارها دارد. همین امر سبب شده که کسب و کارهایی شکل بگیرند که به دنبال سنجش اثرات تصمیم‌ها با استفاده از داده‌ها هستند. لذا فعالیت‌های زیادی از قبیل فروش گزارشات تحلیلی رفتار مشتریان و ذائقه‌سنجی کاربران، پیش‌بینی رویدادها یا رفتار مشتری، کشف تقلب و غیره از جمله مواردی هستند که شرکت‌ها به آنها پرداخته‌اند. از طرفی از آنجا که در فناوری‌های نوظهور از قبیل اینترنت اشیا و غیره، دادگان نقش بسیار زیادی دارد و بصورت جریانی تولید می‌شود بحث کلان داده و پردازش‌های مربوطه از اهمیت فوق‌العاده‌ای برخوردار است. لذا در این مقاله سعی شد تا از طریق یک مدل مفهومی مبتنی بر درختواره به تشریح ابعاد مختلف داده پردازیم. در ادامه تلاش کردیم تا با پیمایش وضعیت استارت‌آپ‌های حوزه داده به استخراج و تحلیل جنبه‌هایی از داده که قابلیت بیشتری برای خلق ارزش و



شکل ۱۲: محصولات حوزه مدیریت داده



شکل ۱۳: محصولات حوزه جمع‌بندی، اشتراک و مصورسازی داده



شکل ۱۴: محصولات حوزه تحلیل محتوا

از آنجائی که محصولات در سال‌های مختلفی تولید و ارائه شده‌اند، لذا نمی‌توان رقم فروش محصولات را با یکدیگر مقایسه کرد. اما با توجه به توزیع محصولات هر زیردسته در سال‌های مختلف و برای اینکه تصویری نسبی از اقبال بازارهای داخل کشور به محصولات مذکور ایجاد شود، در جدول ۱، مقایسه‌ای از فروش محصولات زیردسته‌های مختلف آورده شده است. البته با توجه به توضیحاتی که در بالا ذکر شده است، بجای ذکر ارزش ریالی، به سهم هر دسته از مجموع فروش محصولات مورد بررسی اکتفا شده است.

جدول ۱: مقایسه تعداد و سهم فروش محصولات زیردسته‌های مختلف

سهم فروش از کل	تعداد محصولات	
۰,۳۲٪	۱	پلتفرم‌های اشتراک‌گذاری و

- [۷] محمدشهرام معین، برزگر، نکيسا، وحيد رادمهر، کاربردهای کلان داده، پروژه تدوین نقشه راه کلان داده، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۵
- [۸] محمدشهرام معین، ملکی، داود، محمدرضا احمدی و سیدمحمدرضا میرصراف، ذخیره‌سازی داده‌ها، پروژه تدوین نقشه راه کلان داده، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۵
- [۹] محمدشهرام معین، جم‌پور، مهدی، نکيسا برزگر، اکبر دارابی و سمیه فتاحی، تحلیل کلان‌داده‌ها، پروژه تدوین نقشه راه کلان داده، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۵
- [۱۰] A. Siddiqua, I. A. T. Hashem, I. Yaqoob, et al., "A survey of big data management: Taxonomy and state-of-the-art", *Journal of Network and Computer Applications*, vol. ۷۱, pp. ۱۵۱-۱۶۶, ۲۰۱۶
- [۱۱] محمودی، مریم، همایون بهشتی و علیرضا یاری، مدل مفهومی موتور جستجوی فارسی، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۸۹
- [۱۲] محمودی، مریم، مژگان فرهودی، علی شریفی، علیرضا یاری و کامبیز بدیع، ارائه مدل مفهومی پیشنهادی جویس در فضای وب، اولین همایش ملی جویسگر بومی، ۱۳۹۴
- [۱۳] فرهودی، مژگان، شهریار سموری و علیرضا یاری، مدل مفهومی سامانه پرسش و پاسخ، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۸۹
- [۱۴] عظیم‌زاده، معصومه، محمد امین فرجیان و علیرضا یاری، مدل مفهومی ترجمه ماشینی با هدف تهیه قالب‌های تعریف پروژه‌ها و تدوین برنامه اجرایی آن، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۸۹
- [۱۵] محمدشهرام معین، ملکی، داود، محمدرضا احمدی و سیدمحمدرضا میرصراف، ذخیره‌سازی داده‌ها، پروژه تدوین نقشه راه کلان داده، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۵
- [۱۶] R. Vatananan-Thesenvitz, N. Gerdri. "The current state of technology roadmapping (TRM) research and practice", *International Journal of Innovation and Technology Management*, vol. ۹, no. ۴, pp. ۱-۱۰, ۲۰۱۰
- [۱۷] یاری، علیرضا، احسان آریانیان و مژگان فرهودی، "ارائه مدل مفهومی روندها و حوزه‌های تحقیقاتی داده‌های عظیم"، همایش داده‌های عظیم، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۳
- [۱۸] S. Karamizadeh, Sh. M. Abdullah, M. Zamani, A. Kheirkhah, "Pattern Recognition Techniques: Studies on Appropriate Classification", *ARNP*
- درآمد دارند، برسیم. نتایج بررسی‌ها نشان داد که دو حوزه تحلیل و مصورسازی داده بیش از دیگر حوزه‌ها در داخل کشور مورد توجه قرار گرفته است که دلیل آن هم کاربرد و نیز ارزش‌آفرینی بیشتر این دو حوزه در صنایع و کسب‌وکارهای مختلف است. همچنین نتایج نشان داده که جدا از کاربردهای عمومی دو حوزه «فین‌تک» و «رسانه» بیش از سایرین مورد توجه بوده است. همچنین پس از بررسی سهم فروش محصولات این دسته در کشور، مشخص شد که محصولات مربوط به «مدیریت و پردازش کلان‌داده‌ها» که شامل ارائه راهکار و راه‌اندازی خدمات پردازش و ذخیره‌سازی کلان داده‌ها می‌شود، بیشترین حجم فروش را به خود اختصاص داده است. لازم به ذکر است که میزان قابل توجهی از شرکت‌های مورد بررسی بر روی محتوای اخبار، شبکه‌های اجتماعی و اطلاعات مربوط به معاملات بورس سرمایه‌گذاری کرده‌اند. البته دلیل وجود داده‌های کلان و با ارزش در اغلب شرکت‌ها، نهادها و سازمان‌های مختلف هنوز جای کار زیادی در این حوزه وجود دارد که می‌توان با درک اهمیت آن به ارزش‌آفرینی بیشتری در حوزه‌ها و صنایع مختلف از قبیل گردشگری، سلامت، کشاورزی، بانکداری و غیره پرداخت.

مراجع

- [۱] D. Reinsel, J. Gantz, J. Rydning, "The Digitization of the World from Edge to Core", IDC White Paper, ۲۰۱۸
- [۲] European Council Conclusion, Building a Data Economy in the European Union Data. Innovation. Growth., ۲۰۱۹
- [۳] B. M. Balachandran, Sh. Prasad, "Challenges and Benefits of Deploying Big Data Analytics in Cloud for Business Intelligence", *Procedia Computer Science*, vol. ۱۱۲, pp. ۱۱۱۲-۱۱۲۲, ۲۰۱۷
- [۴] Ph. Hartmann, M. Zaki, N. Feldmann, A. Neely, "Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms", University of Cambridge, ۲۰۱۴
- [۵] E. Park, L. Chung, H. M. Johng, V. Sugumaran, et al., "A Big Data Conceptual Model to Improve Quality of Business Analytics", in *Research Challenges in Information Science*, pp ۲۰-۳۷, June ۲۰۲۰
- [۶] M. Halaweh, A. E. Massry, "Conceptual Model for Successful Implementation of Big Data in Organizations", *Journal of International Technology and Information Management*, vol. ۲۴, Issue ۲, pp. ۲۱-۳۴, ۲۰۱۵

Journal of Engineering and Applied Sciences,
vol. ۹, no. ۸, pp. ۱۲۱۵-۱۲۱۹, August ۲۰۱۴

[۱۹] P. Maroufkhani, R. Wagner, W. Kh. W. Ismail, M. Nourani, "Big Data Analytics and Firm Performance: A Systematic Review", Journal of Information, vol. ۱۰, No. ۲۲۶, pp. ۱-۲۱, July ۲۰۱۹

[۲۰] سخایی، محمدجواد، تجزیه و تحلیل داده و انواع آن، فناوری اطلاعات برای کسب و کار، ۱۳۹۴

[۲۱] درودی، فریبرز، عادل سلیمانی‌نژاد، "ارتباط تصویری و پردازش دیداری اطلاعات: آشنایی با مبانی نظری حوزه مصورسازی"، مجله مطالعات ایرانی کتابداری و اطلاع‌رسانی، شماره ۸۶، ۱۳۸۸

[۲۲] بنائی، سیدمجتبی، دریاچه داده به عنوان بستر حکمرانی داده در سازمان، چارچوب‌های پردازش کلان داده، ۱۳۹۶

سپاسگزاری

بدینوسیله از کارگروه محترم ارزیابی و تشخیص صلاحیت شرکت‌ها و موسسات دانش بنیان که ما را در انجام این تحقیق یاری نمودند صمیمانه تشکر می‌نمایم.