

اقتصاد و کلان داده‌ها:

**کلان داده‌ها چگونه ارزش‌های
جدید اقتصادی خلق می‌کنند؟**



معاونت بررسی‌های اقتصادی
اتاق بازرگانی، صنایع، معادن و کشاورزی تهران





تهیه و تنظیم: محمد رهبری

معاونت بررسی های اقتصادی

اتاق بازرگانی، صنایع، معادن و کشاورزی تهران

آبان ۱۴۰۲

استفاده از مطالب این گزارش با ذکر منبع بلامانع است.



فهرست مطالب

۴	خلاصه مدیریتی
۵	مقدمه
۶	کلان داده چیست؟
۷	بهره برداری از کلان داده ها
۱۰	کلان داده برای سیاستگذاری و تحلیل اقتصادی
۱۱	کلان داده ها در اقتصاد
۱۱	داده های اداری
۱۲	داده های مالی
۱۳	بازارهای کار
۱۴	داده های متنی
۱۵	داده های تلفن همراه
۱۵	داده های اینترنتی
۱۶	داده های دیگر
۱۷	زنجیره ارزش کلان داده
۱۹	اکوسیستم کلان داده
۲۰	نمونه موردی: اکوسیستم کلان داده اروپایی
۲۲	نتیجه گیری

خلاصه مدیریتی

حجم داده‌ها به طور تصاعدی در حال افزایش است و انتظار می‌رود تا سال ۲۰۲۵ بیش از ۱۸۱ زتابایت^۱ (۱۸۱ تریلیون گیگابایت) داده مفید وجود داشته باشد. ما در آستانه عصری هستیم که هر دستگاهی آنلاین است، عصری که در آن حسگرها در جهان ما همه جا وجود دارند و جریان‌های پیوسته‌ای از داده‌ها را تولید می‌کنند، عصری که حجم انبوه داده‌های ارائه شده و مصرف شده در اینترنت به صورت تصاعدی افزایش می‌یابد، عصری که اینترنت اشیا (Internet of Things) اثر انگشت دیجیتالی از دنیای ما تولید خواهند کرد.

به‌طور خلاصه کلان‌داده را می‌توان با سه ویژگی ۱- حجم بسیار زیاد ۲- سرعت تولید زیاد- تنوع داده (اعم از متنی، تصویری، عددی و غیره)، توصیف کرد. فرصت‌های جدید برای ایجاد مجموعه داده‌های جدید از اطلاعات بدون ساختار قبلی، مانند متن و تصاویر ماهواره‌ای، فراوان است. این توسعه زمینه‌های جدیدی از مطالعات اقتصادی را باز کرده است. سؤالاتی که قبلاً فقط چندین ماه یا حتی سال‌ها پس از وقوع قابل پاسخگویی بودند، اکنون در همان زمان می‌توان به آنها پرداخت. بنابراین، اقتصاددانان از پیش‌بینی به سمت مطالعه حرکت کرده‌اند.

همه این مسائل باعث شده تا مطالعه کلان‌داده‌ها و نقش آن در اقتصاد و سیاست‌گذاری اقتصادی اهمیت روزافزونی پیدا کند. اکنون بسیاری از کسب و کارها در یک جامعه پلتفرمی می‌تواند حول همین کلان‌داده‌ها شکل بگیرد. به‌همین جهت شناخت آن، ارزشی که از طریق آن ایجاد می‌شود و اکوسیستمی که حول آن شکل می‌گیرد، اهمیت دوچندانی دارد.

واقعیت آن است کلان‌داده حوزه نوظهوری است که در آن فناوری‌های خلاقانه، راه‌های جدیدی برای استخراج ارزش از موج اطلاعات جدید ارائه می‌دهد. از کلان‌داده‌ها می‌توان در حوزه‌های مراقبت‌های بهداشتی، بخش عمومی، امور مالی و بیمه، رسانه و سرگرمی و خرده‌فروشی بهره‌برداری کرد. انواع داده‌های قابل بهره‌برداری در این حوزه داده‌های اداری، داده‌های مالی، بازار کار، داده‌های متنی و داده‌های اینترنتی هستند.

زنجیره ارزش کلان‌داده این فعالیت‌های کلیدی سطح بالا را شناسایی می‌کند: اکتساب داده، تحلیل داده، ساماندهی داده، ذخیره‌سازی داده و استفاده از داده.

اکوسیستم‌های کلان‌داده در بسیاری از بخش‌های صنعتی وجود دارند که در آن حجم وسیعی از داده‌ها بین بازیگران زنجیره‌های تامین اطلاعات پیچیده جابجا می‌شود. بخش‌های تجاری حاوی اکوسیستم‌های داده تثبیت‌شده یا نوظهور عبارتند از مراقبت‌های بهداشتی، مالی، لجستیک، رسانه، تولید و داروسازی. علاوه بر خود داده‌ها، اکوسیستم‌های کلان‌داده می‌توانند توسط پلتفرم‌های مدیریت داده، زیرساخت‌های داده (مانند پروژه‌های متن‌باز مختلف آپاچی) و سرویس‌های داده نیز پشتیبانی شوند.

^۱ Zettabyte

مقدمه

عصر فناوری که اکنون در آن زندگی می کنیم را می توان عصر «تحول دیجیتال» نامید. رایانه ها، تلفن ها و بسیاری از فناوری های دیگر ارزان، قابل حمل و سریع شده اند، و در نتیجه در همه جا وجود دارند. امروزه همه چیز به سرعت در حال تغییر است. محرک اصلی این تغییرات، مصرف کننده «دیجیتال» است. سرعت تغییر در ترجیحات مشتری، باعث افزایش تقاضا برای کیفیت، عملکرد و طراحی محصولات شده است.

احساسات اکنون برای مصرف کننده مهم هستند. الگوهای مصرف در حال تغییر هستند. این تغییر از روندهای آشکار موجود پدیدار می شود: تعامل از طریق دستگاه های هوشمند، عملکرد کسب و کار با سرعت بالا، افزایش تقاضا، اشباع بازارهای سنتی برای شرکت ها، افزایش رقابت و ظهور رقبا فناورانه جدید. فناوری های منسوخ برای مدت طولانی سود زیادی به ارمغان نمی آورند، مدرن سازی آن ها گران است و به منابع نسبتاً زیادی برای نگهداری آنها نیاز دارد.

بازار جهانی کلان داده ها بسیار سریع در حال رشد است. فعال ترین تولیدکنندگان و مصرف کنندگان داده، بخش عمومی، مراقبت های بهداشتی، تولید و خرده فروشی هستند. با این حال، حجم داده ها به طور تصاعدی در حال افزایش است و انتظار می رود تا سال ۲۰۲۵ بیش از ۱۸۱ زتابایت (۱۸۱ تریلیون گیگابایت) داده مفید وجود داشته باشد. ما در آستانه عصری هستیم که هر دستگاهی آنلاین است، عصری که در آن حسگرها در جهان ما همه جا وجود دارند و جریان های پیوسته ای از داده ها را تولید می کنند، عصری که حجم انبوه داده های ارائه شده و مصرف شده در اینترنت با به صورت تصاعدی افزایش می یابد، عصری که اینترنت اشیا (Internet of Things) اثر انگشت دیجیتالی از دنیای ما تولید خواهند کرد.

در حالی که به طور سنتی داده ها فقط برای یک هدف خاص، اغلب توسط یک نهاد ملی آماری، جمع آوری می شدند، اکنون جهان به طور فزاینده ای در حال کمی شدن است، جایی که حتی کوچکترین شرکت ها داده های دقیق و گاهی فردی را جمع آوری و ثبت می کنند. این کار از طریق اکوسیستم وسیعی از نرم افزارها (برنامه ها) و سخت افزار (حسگرها) تعبیه شده در دریای وسیع فناوری «هوشمند» از جمله تلفن ها، وسایل متصل به Wi-Fi، اتومبیل ها و ماهواره ها انجام می شود. این حجم داده ها به طور چشمگیری هم تنوع داده ها و هم سرعت ثبت داده ها را افزایش داده است. فرصت های جدید برای ایجاد مجموعه داده های جدید از اطلاعات بدون ساختار قبلی، مانند متن و تصاویر ماهواره ای، فراوان است. این توسعه زمینه های جدیدی از مطالعات اقتصادی را باز کرده است. سؤالاتی که قبلاً فقط چندین ماه یا حتی سال ها پس از وقوع قابل پاسخگویی بودند، اکنون در همان زمان می توان به آن ها پرداخت. بنابراین، اقتصاددانان از پیش بینی به سمت مطالعه زمان کنونی حرکت کرده اند. به عنوان مثال، اکنون می توان از جستجوهای گوگل در زمان واقعی برای پیش بینی تغییرات بیکاری یا از داده های Yelp برای پیش بینی الگوهای کسب و کار محلی استفاده کرد.

همه این موارد، آشنایی با کلان داده ها و روش های تحلیل آن را برای اقتصاددانان، صاحبان کسب و کار و سیاستگذاران اقتصادی ضروری می کند. با این حال این حوزه در کشور ما هنوز تا حد زیادی مغفول واقع شده است و بسیاری از سیاستگذاران و صاحبان کسب و کار با مفاهیمی اساسی مربوط به این حوزه آشنایی درستی ندارند. این امر اهمیت پرداختن به این موضوع را دوچندان می کند.

واقعیت آن است کلان داده حوزه نوظهوری است که در آن فناوری‌های خلاقانه، راه‌های جدیدی برای استخراج ارزش از موج اطلاعات جدید ارائه می‌دهد. اکنون توانایی مدیریت موثر اطلاعات و استخراج دانش به عنوان یک مزیت رقابتی کلیدی در نظر گرفته می‌شود. بسیاری از سازمان‌ها کسب و کار اصلی خود را بر اساس توانایی جمع‌آوری و تحلیل اطلاعات برای استخراج دانش و بینش تجاری می‌سازند. حالا دیگر پذیرش فناوری کلان داده‌ها در بخش‌های صنعتی یک امر لوکس نیست، بلکه یک نیاز ضروری برای بیشتر سازمان‌ها برای بقا و کسب مزیت رقابتی است و آشنایی با ابعاد مختلف آن از این حیث اهمیت دارد.

کلان داده چیست؟

اصطلاح «کلان داده» در حدود سال ۲۰۱۰ وارد واژگان اصلی شد، زمانی که مردم از سرعت نمایی تولید داده‌ها، عمدتاً از طریق استفاده از رسانه‌های اجتماعی، آگاه شدند. مهندسان و دانشمندان کامپیوتر به سرعت دریافتند که کلان داده را نمی‌توان صرفاً بر اساس اندازه تعریف کرد. در حالی که مطمئناً درست است که حجم داده‌ها در طول دهه گذشته مرتباً افزایش یافته است، عوامل دیگری نیز چشم‌انداز اطلاعاتی را تغییر داده‌اند.

در طول سال‌های گذشته، اصطلاح کلان داده^۲ توسط بازیگران مختلفی برای نام‌گذاری داده‌ها با ویژگی‌های مختلف استفاده شده است. در دهه گذشته تعاریف متعددی از کلان داده ارائه شده است. تعریف اول، توسط داگ لین از گروه متا^۳ (که در آن زمان توسط گارتنر خریداری شد)، کلان داده‌ها را با استفاده از یک چشم‌انداز سه بعدی تعریف کرد: «داده‌های بزرگ حجم بالا، سرعت بالا و یا دارایی‌های اطلاعاتی با تنوع بالا هستند که برای فعال کردن تصمیم‌گیری پیشرفته، کشف بینش و بهینه‌سازی فرآیند به اشکال جدیدی از پردازش نیاز دارند». لوکیدس کلان داده‌ها را اینگونه تعریف می‌کند: «زمانی که اندازه داده به خودی خود بخشی از مشکل می‌شود و تکنیک‌های سنتی کار با داده‌ها بکار نمی‌آید». جاکوبز (۲۰۰۹) کلان داده را به عنوان "داده‌هایی که اندازه آن‌ها ما را مجبور می‌کند فراتر از روش‌های آزمایش شده و واقعی که در آن زمان رایج است نگاه کنیم" توصیف می‌کند.

کلان داده‌ها مجموعه‌ای از چالش‌های مدیریت داده را برای کار با داده‌ها در مقیاس‌های جدید از منظر اندازه و پیچیدگی گرد هم می‌آورد. بسیاری از این چالش‌ها جدید نیستند. با این حال، آنچه جدید است، چالش‌های ناشی از ویژگی‌های خاص داده‌های بزرگ مربوط به ۳ V است (V حرف ابتدایی این سه چالش بزرگ در زبان انگلیسی است):

- حجم^۴ (مقدار داده): به حجم عظیم داده در پردازش داده‌ها اشاره دارد (مانند زنجیره‌های تامین جهانی، تحلیل مالی جهانی، برخورد دهنده هادرون بزرگ).
- سرعت^۵ (سرعت داده): به سرعت سریع تولید داده به صورت لحظه‌ای اشاره دارد (مانند حسگرها، محیط‌های فراگیر، تجارت الکترونیک، شبکه بین ایشیا).

^۲- Big Data

^۳- Meta Group

^۴- Volume

^۵- Velocity

• تنوع⁶ (محدوده ای از انواع داده ها/منابع): به تنوع اشکال مختلف داده اشاره دارد(مانند صفحات گسترده، XML، DBMS)، طرحواره‌ها و معانی (مانند یکپارچه سازی داده‌های سازمانی).

کلان داده‌ها اصول رویکردهای فنی موجود را به چالش می کشد و نیاز به اشکال جدیدی از پردازش داده‌ها را ایجاد می کند تا امکان تصمیم گیری، کشف بینش و بهینه سازی فرآیند افزایش یابد. با بلوغ حوزه کلان داده، چالش‌های دیگری مانند حقیقت (کیفیت داده و عدم قطعیت)، ارزش و غیره اضافه شده است. ارزش کلان داده‌ها را می توان در زمینه پویایی سازمان‌های مبتنی بر دانش توصیف کرد، جایی که فرآیندهای تصمیم گیری و اقدام سازمانی به فرآیند حس سازی و خلق دانش وابسته است.

بهره برداری از کلان داده‌ها

اثرات کلان داده فراتر از دنیای تجاری است. در جامعه علمی، انفجار داده‌های موجود چیزی را ایجاد می کند که علم داده نامیده می شود؛ یک رویکرد جدید مبتنی بر داده برای کشف علمی. توانایی تلسکوپ‌ها یا شتاب دهنده‌های ذرات برای تولید چندین پتابایت داده در روز، مشکلات مختلفی را از نظر ذخیره سازی و پردازش ایجاد می کند. دانشمندان راه حل‌های آماده‌ای برای تحلیل و مقایسه صحیح مجموعه داده‌های پراکنده و عظیم ندارند. فعال کردن این چشم انداز به فناوری‌های نوآورانه کلان داده‌ها برای مدیریت داده، پردازش، تحلیل، کشف و استفاده نیاز دارد.

داده‌ها همانند دارایی‌های سخت و سرمایه انسانی به عامل جدیدی برای تولید تبدیل شده‌اند. داشتن پایه تکنولوژیکی و ساختار سازمانی مناسب برای بهره برداری از داده‌ها ضروری است. اکنون تمام کشورها باید از پتانسیل داده‌های بزرگ برای ایجاد ارزش برای جامعه، شهروندان و کسب و کارها استفاده کنند و استفاده از آن می تواند به یک مزیت رقابتی تبدیل شود. برای افزایش رقابت پذیری صنایع از طریق کلان داده‌ها به یک استراتژی روشن نیاز است و ضرورت دارد که برای آن برنامه ریزی مناسب شود. عدم ورود مناسب به این حوزه، کشورها را وابسته به خارج کرده و ممکن است ذینفعان را از استفاده کامل فناوری کلان داده بازدارد. رقابت در فناوری‌ها و راه حل‌های کلان داده به کشورها، از جمله کشورهای در حال توسعه، منبع جدیدی از رقابت و پتانسیل ایجاد یک صنعت جدید مرتبط با داده را می دهد که مشاغل جدید ایجاد می کند.

تحول بخش‌های صنعتی

پیش بینی می شود فواید بالقوه کلان داده‌ها بر همه بخش‌ها، از جمله مراقبت‌های بهداشتی، رسانه‌ها، انرژی و خرده فروشی تأثیر بگذارد. پتانسیل تحول مثبت قبلاً در تعدادی از بخش‌های کلیدی شناسایی شده است.

⁶ Variety

مراقبت‌های بهداشتی: اکنون بسیاری از جوامع به سمت سالخورده‌گی حرکت می‌کنند و تا چند دهه آینده، بحران جمعیت سالخورده یکی از بحران‌های عمومی در بسیاری از کشورها، از جمله ایران، خواهد بود؛ همین امر باعث افزایش مطالبات برای توسعه زیرساخت‌های مراقبت‌های بهداشتی می‌شود. به همین دلیل نیاز فوری به توسعه و بهبود کارایی سیستم مراقبت‌های بهداشتی فعلی برای پایدارتر کردن آن وجود دارد. از آنجایی که هزینه توسعه و بهبود این سیستم‌ها می‌تواند بالا باشد، استفاده از کلان‌داده‌ها یک مزیت مهم برای این کار ایجاد می‌کنند و می‌توانند باعث صرفه‌جویی در هزینه‌های این کار شوند. به طور مثال، در اتحادیه اروپا، استفاده از کلان‌داده‌ها می‌تواند منجر به صرفه‌جویی ۹۰ میلیارد یورویی در هزینه‌های مربوط به این بخش شود. کاربردهای بالینی کلان‌داده گسترده وسیعی را شامل می‌شود؛ از جمله تحقیقات راجع به اثربخشی بالینی و مالی مداخلات تا نسل بعدی سیستم‌های پشتیبانی تصمیم بالینی که از مجموعه داده‌های بهداشتی ناهمگن جامع و همچنین تحلیل‌های پیشرفته عملیات بالینی استفاده می‌کنند. برنامه‌های تحقیق و توسعه مراقبت‌های بهداشتی شامل مدل‌سازی پیش‌بینی‌کننده، ابزارهای آماری و الگوریتم‌هایی برای بهبود طراحی کارآزمایی بالینی، پزشکی شخصی سازی شده و تحلیل الگوهای بیماری‌ها می‌شود.

بخش عمومی: بخش عمومی بخش مهمی از تولید ناخالص داخلی کشورها را تشکیل می‌دهد و همچنین می‌تواند به طور قابل توجهی از کلان‌داده‌ها برای به دست آوردن کارایی در فرآیندهای اداری بهره‌مند شود. پیش‌بینی می‌شود که کلان‌داده‌ها می‌توانند هزینه‌های فعالیت‌های اداری را به میزان ۱۵ تا ۲۰ درصد کاهش دهند. مزایای بالقوه بکارگیری کلان‌داده‌ها در بخش عمومی شامل بهبود شفافیت از طریق داده‌های باز (open data)، بهبود تدارکات عمومی، افزایش تخصیص بودجه به برنامه‌ها، خدمات با کیفیت بالاتر، افزایش پاسخگویی بخش عمومی و شهروندی با آگاهی بهتر است. برای آینده، لازم است سیاست‌هایی راجع به اشتراک گذاشتن داده‌ها در میان سازمان‌های دولتی و اطلاع‌رسانی به شهروندان در مورد مخاطرات بین مخاطرات حریم خصوصی و امنیتی ناشی از اشتراک گذاری داده‌ها و مزایایی که می‌توانند به دست آورند، تعریف شود. کلان‌داده‌ها همچنین با توانمندسازی شهروندان برای درک مسائل سیاسی و اجتماعی به روش‌های شفاف جدید، رابطه بین شهروندان و دولت را تغییر می‌دهد و آنها را قادر می‌سازد تا از طریق مشارکت با مسائل محلی، منطقه‌ای، ملی و جهانی درگیر شوند.

امور مالی و بیمه: راه‌های مختلفی برای شرکت‌های خدمات مالی وجود دارد که می‌توانند با استخراج و تحلیل داده‌ها به مزایای تجاری دست یابند. این خدمات شامل افزایش خدمات مشتری خرده‌فروشی، شناسایی تقلب و بهبود کارایی، عملیاتی می‌شود. کلان‌داده‌ها را می‌توان در طیف وسیعی از ابزارهای مالی پیچیده مانند مشتقات مالی^۷ استفاده کرد. تحلیل پیش‌بینی‌کننده داده‌های داخلی و خارجی منجر به مدیریت بهتر و پیشگیرانه طیف گسترده‌ای از مسائل، از ریسک اعتباری و عملیاتی (مانند تقلب و ریسک شهرت) گرفته تا وفاداری و سودآوری مشتری، می‌شود. یک چالش برای بخش مالی این است که چگونه از وسعت و عمق داده‌های موجود برای جلب رضایت طرفداران تنظیم‌گری بیشتر و همچنین ارائه خدمات شخصی برای مشتریان خود استفاده کند.

⁷- derivatives



تلکام^۸، رسانه و سرگرمی: تکنیک های تحلیل و مصورسازی کلان داده ها می تواند ارائه محتوای رسانه ای مؤثر را امکان پذیر کند و کاربران را قادر می سازد تا به طور پویا با رسانه ها و محتوای جدید در چندین پلتفرم تعامل داشته باشند.

خرده فروشی: فرصت های قابل توجهی برای استفاده از فناوری کلان داده در تعاملات بین خرده فروشان و مصرف کنندگان وجود دارد. داده ها در اینجا نقش فزاینده ای ایفا می کنند زیرا مصرف کنندگان به جستجو، تحقیق، مقایسه، خرید و پشتیبانی آنلاین می پردازند و محصولات فروخته شده توسط خرده فروشان به طور فزاینده ای ردپای داده ای خود را ایجاد می کنند. کلان داده ها می توانند بهره وری و کارایی را افزایش دهند که منجر به افزایش بالقوه ۶۰ درصدی حاشیه عملیاتی خرده فروشان می شود. کلان داده ها می توانند بر خرده فروشی در زمینه های زیر تأثیر بگذارند:

- ✓ بازاریابی: فروش متقابل، بازاریابی مبتنی بر مکان، تحلیل رفتار مشتریان، تقسیم بندی خرد مشتری، تحلیل احساسات مشتری، بهبود تجربه مصرف کننده چند کانالی.
- ✓ تجاری سازی: بهینه سازی مجموعه، بهینه سازی قیمت گذاری، بهینه سازی مکان و طراحی.
- ✓ عملیات: شفافیت عملکرد، بهینه سازی ورودی های نیروی کار.
- ✓ زنجیره تامین: مدیریت موجودی، توزیع و بهینه سازی لجستیک، اطلاع رسانی مذاکرات تامین کننده؛ مدل های کسب و کار جدید: خدمات مقایسه قیمت، بازارهای مبتنی بر وب.

تولید: بخش تولید اولین پذیرنده فناوری اطلاعات برای طراحی، ساخت و توزیع محصولات است. نسل بعدی کارخانه های هوشمند با ماشین آلات هوشمند و شبکه ای (یعنی اینترنت اشیا، صنعت ۴.۰) شاهد بهبود کارایی بیشتر در طراحی، تولید و کیفیت محصول خواهد بود. کلان داده ها از طریق محصولات دقیقاً هدفمند و توزیع مؤثر، نیازهای مشتری را برآورده می کند. علاوه بر افزایش بهره وری و نگهداری پیش بینی شده، کلان داده ها مدل های تجاری کاملاً جدیدی را در زمینه تولید انبوه محصولات فردی ایجاد می کند.

انرژی و حمل و نقل: کلان داده ها فرصت های جدیدی را با روش های نوآورانه برای نظارت و کنترل شبکه های حمل و نقل و لجستیک با استفاده از انواع منابع داده و اینترنت اشیا باز می کند. پتانسیل کلان داده ها در بخش حمل و نقل در سراسر جهان در قالب صرفه جویی در زمان و سوخت، با اجتناب از انتشار ۳۸۰ مگاتن CO₂، ۵۰۰ میلیارد دلار در سراسر جهان برآورد شده است. دیجیتالی شدن سیستم های انرژی امکان دستیابی به داده های با وضوح بالا و در زمان واقعی را از طریق کنتورهای هوشمند فراهم می کند که می توانند در تحلیل های پیشرفته برای بهبود سطوح کارایی در هر دو طرف تقاضا و عرضه شبکه های انرژی مورد استفاده قرار گیرند. ساختمان های هوشمند و شهرهای هوشمند محرک های کلیدی افزایش بهره وری در بخش های انرژی خواهند بود. فناوری کلان داده در بخش های آب و برق این پتانسیل را دارد که انتشار CO₂ را تا بیش از ۲ گیگاتن، معادل ۷۹ میلیارد یورو کاهش دهد.

^۸ صنعت تلکام (Telecom) یا صنایع ارتباطات، شاخه ای از فناوری اطلاعات و ارتباطات است، که از کلیه شرکت های ارتباطات، مخابرات و نیز ارائه دهندگان خدمات اینترنتی تشکیل می شود. این صنعت امروزه نقشی اساسی در توسعه ارتباطات سیار و تکامل جوامع اطلاعاتی ایفا می کند.

کلان داده برای سیاست‌گذاری و تحلیل اقتصادی

در دسترس بودن کلان داده‌ها امکان بررسی سوالات طولانی مدت را با استفاده از اطلاعات دقیق‌تر در مورد رفتار اقتصادی فراهم می‌کند. علاوه بر این، این مجموعه داده‌ها امکان کشف حقایق تجربی جدیدی را فراهم می‌کند که قبلاً به دلیل کمبود اطلاعات شناخته شده نبودند.

نقش «کلان داده‌ها» در زمینه برنامه‌های اقتصادی دقیقاً چیست؟ کلان داده‌ها را می‌توان به عنوان مجموعه داده‌هایی تعریف کرد که تحلیلشان به سخت افزار محاسباتی و/یا ابزارهای نرم افزاری پیشرفته نیاز دارند. یکی از این ابزارها محاسبات توزیع شده است که پردازش یک کار را در چندین ماشین به اشتراک می‌گذارد، به جای یک ماشین واحد که معمولاً توسط اقتصاددانان انجام می‌شود. نمونه‌هایی از کلان داده‌های مورد استفاده در تحلیل اقتصادی، داده‌های اداری (مانند سوابق مالیاتی برای کل جمعیت یک کشور)، مجموعه داده‌های تجاری (مانند پانل‌های مصرف‌کننده) و داده‌های متنی (مانند داده‌های توییتر یا اخبار) هستند. در برخی موارد، مجموعه داده‌ها ساختار یافته و آماده برای تحلیل هستند، در حالی که در موارد دیگر (به عنوان مثال، متن)، داده‌ها ساختاری ندارند و نیاز به یک مرحله مقدماتی برای استخراج و سازماندهی اطلاعات مربوطه دارند. اقتصاددانان هنوز در مراحل اولیه تحلیل کلان داده‌ها هستند و از تحولات سایر رشته‌ها درس می‌گیرند. به طور خاص، پس از کاربردهای اولیه در دهه ۱۹۹۰، علاقه مجددی به الگوریتم‌های یادگیری ماشین (ML) وجود دارد.

چگونه کلان داده‌ها می‌توانند به درک بهتر اقتصاد و حمایت از سیاست‌گذاری کمک کنند؟ در سطح بسیار کلی، کلان داده‌ها فرصتی را برای آشکار کردن ناهمگونی در مصرف‌کنندگان و شرکت‌ها فراهم می‌کند که معمولاً در آمارهای رسمی نادیده گرفته می‌شود. جزئیات زیاد کلان داده‌ها می‌تواند برای ساختن شاخص‌هایی که برای توضیح پدیده‌های خاص طراحی شده‌اند، به عنوان مثال، در امتداد یک بعد جغرافیایی یا جمعیتی، مورد سوء استفاده قرار گیرد. علاوه بر این، بسیاری از مدل‌های اقتصادی مفروضاتی در مورد پارامترهای رفتاری عمیق ایجاد می‌کنند که برآورد آنها بدون مجموعه داده‌های دقیق دشوار است. یک مثال در پژوهش چتی و همکاران (۲۰۱۴) نشان داده شده است که در آن اطلاعات فردی در مورد عملکرد مدرسه یک کودک با مسیر درآمدهای آتی او که از داده‌های مالیاتی سرویس درآمد داخلی (IRS) به دست می‌آید، تطبیق داده می‌شود. در موقعیت‌های دیگر، کلان داده‌ها به اندازه‌گیری مقادیری اجازه می‌دهند که تاکنون نمی‌توانستیم اندازه‌گیری کنیم. حوزه‌ای که از این منابع جایگزین داده سود می‌برد، اقتصاد توسعه است. به عنوان مثال، استوریگارد (۲۰۱۶) از داده‌های ماهواره‌ای نور شب برای تخمین درآمد شهرهای جنوب صحرای آفریقا استفاده می‌کند.

یکی دیگر از ابعاد مهمی که در آن کلان داده‌ها می‌توانند تحلیل اقتصادی کمک کنند، ارائه اطلاعاتی است که نه تنها جزئی‌تر هستند، بلکه در بعد زمانی نیز فراوان‌تر هستند. در مواقعی که شرایط اقتصادی به سرعت در حال تغییر است، سیاست‌گذاران برای طراحی پاسخ مناسب سیاستی به اندازه‌گیری دقیق وضعیت اقتصادی نیاز دارند. یک مثال در روزهای اولیه همه‌گیری کووید-۱۹ در مارس ۲۰۲۰ ارائه می‌شود، درست زمانی که سیاست‌گذاران فشار را احساس کردند تا با وجود فقدان آمار رسمی برای اندازه‌گیری میزان رکود، اقدام به حمایت از اقتصاد کنند. کلان داده‌ها کمک می‌کنند تا این داده‌ها زودتر در اختیار اقتصاددانان و سیاست‌گذاران قرار گیرد. بسیاری از

شاخص‌های اقتصادی مرتبط به ندرت مشاهده می‌شوند، مانند تولید ناخالص داخلی (GDP) در دفعات سه ماهه و نرخ بیکاری و شاخص تولید صنعتی در دفعات ماهانه. علاوه بر این، این متغیرها با تاخیرهایی از چند روز تا چند ماه منتشر می‌شوند. به این دلایل، کلان‌داده‌ها پتانسیل تولید شاخص‌هایی از شرایط کسب و کار را دارند که دقیق‌تر و به موقع‌تر باشند.

به طور کلی، شرکت‌های خصوصی در حال جمع‌آوری مقادیر قابل توجهی از داده‌ها هستند که می‌تواند برای تکمیل آمار رسمی و اطلاع‌رسانی سیاست‌های اقتصادی مورد استفاده قرار گیرد. همانطور که توسط Bostic و همکاران بحث شده است. (۲۰۱۶)، رویکرد سازمان‌های دولتی برای تولید آمار رسمی، تا حد زیادی بر اساس نظرسنجی‌های مصرف‌کننده و تجاری است. این رویکرد دقت و نماینده بودن نمونه را تضمین می‌کند، اگرچه روشی گران و زمان‌بر است. از این رو، در دسترس بودن مجموعه داده‌های جایگزین امکان استخراج اطلاعاتی را فراهم می‌کند که می‌تواند شواهد به دست آمده از بررسی‌ها را تکمیل کند.^۹

کلان‌داده‌ها در اقتصاد

در این بخش در مورد یافته‌های اصلی کاربردهای اخیر داده‌های بزرگ در اقتصاد و امور مالی بحث می‌شود. بحث بر اساس منبع داده سازماندهی شده است تا بتوان مروری منسجم‌تر از نتایج ارائه کرد. هدف این بخش جامع نیست، بلکه ارائه یک مرور مختصر از برخی از کاربردهای اصلی داده‌های بزرگ در اقتصاد است.

داده‌های اداری

داده‌های اداری به داده‌های جمع‌آوری شده توسط سازمان‌های دولتی به عنوان بخشی از وظایف آنها اشاره دارد. همانطور که توسط Card و همکاران بحث شده است. (۲۰۱۰)، مزایای اصلی داده‌های اداری، نسبت به نظرسنجی‌ها، نمونه‌های بزرگ آنها، نرخ فرسایش پایین و همچنین خطای کم محاسبه است. علاوه بر این، مجموعه داده‌های اداری از نظر اطلاعات موجود در مورد افراد بسیار دقیق است. با این حال، محقق با توجه به دسترسی محدود به داده‌ها با چالش‌های قابل توجهی در انجام تحلیل مواجه است. به طور معمول، محقق ملزم به ارائه کد به سازمان دولتی است که در واقع تحلیل را انجام می‌دهد و به طور قابل توجهی سرعت توسعه پروژه تحقیقاتی را کاهش می‌دهد.

یک مقاله تأثیرگذار که از داده‌های سوابق مالیاتی استفاده می‌کند، مقاله Chetty و همکاران (۲۰۱۴) است. هدف این مقاله بررسی تحرک بین نسلی در ایالات متحده است. آنها از نمونه‌ای مشکل از ۴۰ میلیون کودک متولد شده بین سال‌های ۱۹۸۰ تا ۱۹۸۲ استفاده می‌کنند و درآمد آنها در سن ۳۰ سالگی را به درآمد والدین مرتبط می‌کنند. این مجموعه داده مدیریتی یک

^۹ -تحلیل عمیق‌تر در مورد استفاده از داده‌های ردیابی دیجیتال و داده‌های غیر متعارف در آمار رسمی را می‌توان در Signorelli et al., 2023 یافت.

تنظیم منحصر به فرد برای ارزیابی تحرک بین نسلی را نشان می دهد زیرا نمونه بزرگی را ارائه می دهد که به سال ۱۹۸۰ برمی گردد و اجازه می دهد تا کودکان و والدین را با دقت بسیار بالا به یکدیگر مرتبط کنند.

اطلاعات سازمان تامین اجتماعی (SSA) در مقاله Kopcuk و همکارانش (۲۰۱۰) استفاده شده است. تا نابرابری درآمد و تحرک اجتماعی در ایالات متحده از سال ۱۹۳۷ را بررسی کند. آنها دریافتند که نابرابری تا اوایل دهه ۱۹۵۰ کاهش یافته و از آن زمان به طور پیوسته افزایش یافته است. از نظر تحرک اجتماعی، آنها نشان می دهند که تحرک اجتماعی در طول زمان، از جمله در انتهای بالای توزیع درآمدی، نسبتاً ثابت بوده است.

کلان داده های اداری همچنین برای ارزیابی دستاوردهای آموزشی و اثربخشی تدریس استفاده می شود. دابی و فرایر (۲۰۱۱) از داده های اداری اداره آموزش و پرورش شهر نیویورک برای ارزیابی تأثیر برنامه های مدارس نمونه دولتی^{۱۰} بر پیشرفت دانش آموزان استفاده می کند. شواهد نشان می دهد که مدارس نمونه دولتی در چندین معیار تأثیر مثبت و قابل توجهی بر بهبود عملکرد تحصیلی کودکان فقیر دارند. یکی از توضیح های ممکن برای این پیشرفت ها این است که این مدارس از معلمان با کیفیت بالا استفاده می کنند. موضوع سنجش کیفیت معلمان و تأثیر آنها بر عملکرد دانش آموزان توسط چتی و همکارانش (۲۰۱۴) بررسی شده است. آن ها از نمونه ای متشکل از یک میلیون دانش آموز استفاده می کنند و داده های مربوط به مناطق مدرسه و سوابق مالیاتی را با هم تطبیق می دهند تا تکامل درآمد کودکان نمونه را دنبال کنند. آنها دریافتند که معیارهای ارزش افزوده معلم (VA)، مانند نمرات آزمون دانش آموز، سوگیری قابل توجهی را به عنوان شاخص کیفیت معلم نشان نمی دهند.

منبع دیگر داده های اداری، ثبت اعتبار^{۱۱} است که توسط Jiménez و همکارانش (۲۰۱۴) استفاده می شود تا اثر سیاست پولی بر رفتار وام دهی بانک را ارزیابی کند. این ثبت اعتبار، کلیه وام ها و قراردادهای بین مردم و بخش بانکی یک کشور را ثبت می کند. آنها نشان می دهند که نرخ بهره پایین تر بر افزایش رفتار ریسک پذیری بانک تأثیر می گذارد که منجر به افزایش عرضه اعتبار، به ویژه برای وام گیرندگان پرریسک تر می شود.

داده های مالی

داده های تراکنش های مالی منبع برجسته ای از کلان داده ها در تحلیل اقتصادی است. یک برنامه مطالعاتی اولیه توسط گراس و سولز (۲۰۰۲) ارائه شده است که از نمونه تصادفی ۲۴ هزار حساب کارت اعتباری برای بررسی تأثیر بدهی تغییرات بر محدودیت های اعتباری استفاده کردند. نتایج آن ها نشان می دهد که افراد با استقراض بیشتر به افزایش محدودیت های اعتباری واکنش نشان می دهند، به ویژه برای کسانی که به محدودیت اعتباری نزدیک هستند. یکی دیگر از برنامه های مطالعاتی جدیدتر با استفاده از داده های کارت اعتباری، پژوهش گالاگر و هارتلی (۲۰۱۷) است که از یک نمونه تصادفی ۵ درصدی از افراد دارای سابقه اعتباری استفاده می کنند. آنها از طوفان کاترینا به عنوان یک آزمایش طبیعی استفاده کردند و متوجه شدند که

¹⁰- charter school

¹¹- credit register

بدهی خانوارهایی که در مناطقی که بیشترین آسیب را از سیل داشته‌اند، کاهش زیادی داشته است، امری که عمدتاً به دلیل کاهش تعهدات وام مسکن است. هوروات و همکاران (۲۰۲۱) از داده‌های کارت اعتباری برای ارزیابی رفتار مصرف‌کنندگان در طول همه‌گیری کرونا در سال ۲۰۲۰ استفاده کردند. آنها دریافتند که مخارج و موجودی کارت اعتباری در ماه مارس/آوریل ۲۰۲۰ به سرعت کاهش یافته است، به ویژه در مناطقی که بیشترین موارد ابتلا را داشتند. بهبود هزینه‌ها در ماه می ۲۰۲۰ آغاز شد و وام‌گیرندگان پرریسک‌تر نسبت به وام‌گیرندگان با امتیاز اعتباری بالا پیش‌تاز بودند. دان و همکاران (۲۰۲۰) از داده‌های کارت اعتباری روزانه برای ارزیابی تأثیر جغرافیایی و بخشی این بیماری همه‌گیر بر هزینه‌های مصرف‌کننده استفاده کردند. آن‌ها نشان دادند که معیار هزینه‌های آن‌ها نزدیک به آمار رسمی ماهانه خرده‌فروشی است، که نشان‌دهنده فواید استفاده از کلان داده برای نظارت بر اقتصاد در زمان واقعی (real-time) است.

کالوت و همکاران (۲۰۰۹) از داده‌های اداری در مورد دارایی‌ها و اطلاعات جمعیتی همه مالیات‌دهندگان در سوند استفاده کردند. هدف این مقاله ارزیابی پیچیدگی مالی خانوارها در اجتناب از اشتباهات سرمایه‌گذاری مانند تنوع کم، اینرسی در ریسک‌پذیری و نگهداری سهام زیانده در حین فروش سهام برنده است. آنها دریافتند که خانوارهایی با سطح ثروت و تحصیلات بالاتر پیچیده‌تر عمل می‌کنند و کمتر مستعد اشتباهات سرمایه‌گذاری هستند.

بازارهای کار

آمارهای بازار کار به دلیل دخالت مستقیم سازمان‌های دولتی در اداره مزایای بیکاری، از لحاظ تاریخی سرشار از داده بوده است. اخیراً شرکت‌های خصوصی شروع به جمع‌آوری اطلاعات در مورد بازار کار کرده‌اند. طبیعتاً سؤال این است که آیا این مجموعه داده‌های خصوصی می‌تواند کل بازار کار و اقتصاد ایالات متحده را نمایندگی کند یا خیر. هورتون و تامب (۲۰۱۵) بحث مفصلی در مورد مزایا و معایب استفاده از چنین داده‌هایی ارائه می‌دهند.

نمونه‌ای از استفاده داده‌های جایگزین بازار کار برای سیاست‌گذاری توسط کایتر و همکاران (۲۰۱۹) ارائه شده است. آنها از داده‌های حقوق و دستمزد شرکت خصوصی ADP برای ایجاد اقدامات استخدامی مشابه آنچه توسط اداره آمار کار (BLS) با استفاده از آمار اشتغال فعلی (CES) ساخته شده است، استفاده می‌کنند. آنها دریافتند که دو معیار اشتغال مکمل یکدیگر هستند و به طور مشترک اطلاعاتی در مورد پویایی بازار کار ارائه می‌دهند. این کمک بسیار مهمی است زیرا نشان می‌دهد که داده‌های جایگزین می‌توانند اطلاعاتی را ارائه دهند که مکمل و همبستگی بالایی با آمار رسمی دارند. مزیت اضافی این منابع داده خصوصی این است که در فرکانس‌های بالاتر در دسترس هستند و به محقق اجازه می‌دهند نمونه را از نظر جغرافیایی و ویژگی‌های جمعیتی تقسیم کنند. این مزیت در پژوهش Cajner و همکاران (۲۰۲۰) مورد بحث قرار گرفته است که در آن پژوهش رفتار بلادرنگ (real-time) اشتغال هفتگی را در طول همه‌گیری کووید-۱۹ نسبت به آمار رسمی ماهانه CES نشان می‌دهد. نتایج مشابهی نیز توسط گریگوری و ژو (۲۰۱۴) به دست آمده است.

داده‌های متنی

یک منبع جایگزین داده که در اقتصاد و امور مالی مورد توجه قرار می‌گیرد، داده‌های متنی است. در این مورد، هدف استفاده از متن روزنامه‌ها، سخنرانی‌ها، گزارش‌های شرکتی و توییت، ساختن معیارهایی است که به درک رفتار اقتصادی یا پیش‌بینی متغیرهای اقتصادی کمک می‌کند. گنتز کوو و همکاران (۲۰۱۹) یک نمای کلی از کارهای انجام شده تاکنون ارائه داده‌اند.

یک منبع مهم از داده‌های متنی، مقالات روزنامه‌ها است که ممکن است به عنوان نماینده‌ای برای مجموعه اطلاعات در دسترس عموم در هنگام تصمیم‌گیری اقتصادی در نظر گرفته شود. مقاله اولیه Tetlock (۲۰۰۷) است که احساسات را از ستونی از وال استریت ژورنال استخراج کرد و متوجه شد که تحلیل احساسات این مقالات، برای پیش‌بینی بازده روزانه بازار کل مفید است. بیکر و همکاران (۲۰۱۶) با شمارش تعداد مقالاتی که حاوی مجموعه‌ای از کلمات کلیدی مرتبط با عدم قطعیت هستند، عدم قطعیت اقتصادی و سیاسی را محاسبه کردند. آنها نشان دادند که محاسبه آنها با معیارهای عدم قطعیت همبستگی زیادی دارد. سایر مطالعات اخیر اخبار به‌عنوان نماینده‌ای از احساسات اقتصادی تحلیل می‌کنند^{۱۲}. نظارت بر احساسات مصرف‌کنندگان و کسب و کارها یک پیشینه طولانی در اقتصاد دارد و معمولاً بر اساس نظرسنجی است. این مقالات نشان می‌دهد که تحلیل احساسات مبتنی بر مقالات و یادداشت‌های روزنامه‌ها نتایج مشابه با تحلیل احساسات مبتنی بر نظرسنجی دارد. مشخص شده است که این شاخص‌ها دارای قدرت پیش‌بینی برای چندین متغیر کلان اقتصادی هستند که نسبت به پیش‌بینی کننده‌های معمولی اقتصاد کلان رو به افزایش است. لارسن و ثوررود (۲۰۱۹) در پژوهششان رابطه بین اخبار و انتظارات مصرف‌کننده را بررسی کردند و دریافتند که موضوعات استخراج شده از اخبار به توضیح تصمیم مصرف‌کنندگان برای به روز رسانی انتظارات تورمی‌شان کمک می‌کند.

خط دیگری از تحقیقات به بررسی نقش ارتباطات در اجرای سیاست پولی پرداخته است. هانسن و مک ماهون (۲۰۱۶) از متن ارتباطات شفاهی و نوشتاری توسط فدرال رزرو برای درک نقش آن در پیش‌بینی متغیرهای اقتصادی استفاده کردند. هانسن و همکاران (۲۰۱۸) نقش افزایش شفافیت در ارتباطات بانک مرکزی را با تحلیل مشورت داخلی سیاستگذاران بررسی کردند. آنها متوجه شدند که الگوهای ارتباطی آنها پس از معرفی شفافیت به طور قابل توجهی تغییر کرده است.

پروژه GDELT^{۱۳} منبع دیگری از داده‌های متنی است که در چندین برنامه پژوهشی مورد استفاده قرار گرفته است. کنسولی و همکاران (۲۰۲۱) از تحلیل احساسات برای درک پویایی بازدهی دولت استفاده کردند. Acemoglu و همکاران (۲۰۱۸) از GDELT برای شناسایی رویدادهای ناآرامی سیاسی و اجتماعی در مصر و ارزیابی تأثیر آنها بر بازده سهام استفاده می‌کند. منبع داده دیگری که در تحلیل اقتصادی و مالی به‌مرور بیشتر مورد استفاده قرار می‌گیرد، توییت است. بیکر و همکاران (۲۰۲۱) از پیام‌های توییت برای ایجاد یک شاخص عدم قطعیت اقتصادی توییت (TEU^۴) مشابه شاخص EPU پیشنهاد شده توسط

Barbaglia et al., forthcoming; Larsen & Thorsrud, 2019; Shapiro et al., 2020; Thorsrud, 2020 – ^{۱۲}

^{۱۳}- برای مطالعه بیشتر راجع به GDELT اینجا را مطالعه کنید: [/https://www.gdelproject.org](https://www.gdelproject.org)

^{۱۴}- Twitter Economic Uncertainty

بیکر و همکاران (۲۰۲۱) که بر اساس مقالات روزنامه است، استفاده می‌کند. نتایج آنها نشان می‌دهد که همبستگی بسیار بالایی بین TEU و EPU وجود دارد.

داده‌های تلفن همراه

داده‌های تلفن همراه منبع دیگری از کلان داده‌ها برای تحلیل اقتصادی است. این نوع داده‌ها به طور بالقوه پتانسیل بسیار بالایی دارند زیرا موقعیت مکانی کاربر را در طول زمان ردیابی می‌کنند. یک برنامه پژوهشی اقتصادی توسط Blumenstock و همکارانش (۲۰۱۵) ارائه شده است که از داده‌های تلفن همراه برای محاسبه وضعیت اجتماعی-اقتصادی تماس گیرنده استفاده می‌کنند. این یک ابتکار بسیار مفید برای کشورهای در حال توسعه است که آمار رسمی آنها چندان قابل اعتماد و توسعه نیافته است. میلوخوا (۲۰۲۰) از داده‌های تلفن همراه برای ردیابی تأثیر حرکت افراد از مناطق پر بیماری به مناطق کم بیماری بر شیوع مالاریا استفاده می‌کند. ایده مشابهی در Iacus و همکارانش (۲۰۲۰) توسعه داده شده است که تأثیر اقدامات برای مهار ویروس کووید-۱۹ بررسی می‌کند. یافته‌های آنها نشان می‌دهد که محاسبه تحرک ساخته شده از داده‌های تلفن همراه، پیش‌بینی‌کننده بسیار دقیقی از شیوع اولیه ویروس در ایتالیا و فرانسه است.

داده‌های اینترنتی

ظهور اینترنت این فرصت را برای محققان ایجاد کرده است تا داده‌های آنلاین را به نمایندگی از متغیرهای اقتصادی مورد علاقه جمع‌آوری کنند.^{۱۵} یک مثال با ظهور eBay به عنوان بازاری برای مبادله کالا ارائه شده است که به اقتصاددانان اجازه می‌دهد مکانیسم‌های طراحی بازار را آزمایش کنند و رفتار پیشنهاد دهندگان و فروشندگان را بررسی کنند. یک مقاله ابتدایی در این حوزه، مقاله باجاری و هورتاکسو (۲۰۰۳) است که قوانین تجربی حراج‌های eBay را بررسی کرده و مدلی از مناقصه را تخمین می‌زند.

حوزه‌ای که اخیراً به شدت به آن پرداخته می‌شود، محاسبه روابط اجتماعی افراد بر اساس پلتفرم‌های آنلاین، مانند فیس بوک بوده است. بیلی و همکارانش (۲۰۱۸a) در مقاله‌شان ساخت شاخصی به نام شاخص اتصال اجتماعی^{۱۶} (SCI) را مورد بحث قرار می‌دهند که ارتباطات دوستی بین کاربران فیس بوک ساکن در مناطق مختلف جغرافیایی ایالات متحده و خارج از کشور را اندازه‌گیری می‌کند. کاربرد SCI برای توضیح بازار مسکن در مقاله بیلی و همکاران (۲۰۱۸) ارائه شده است. آنها دریافتند که ارتباطات اجتماعی به توضیح افزایش قیمت مسکن کمک می‌کند که به عقیده آنها نتیجه مشابهت تجربه و انتظارات در مورد بازار مسکن است.

^{۱۵} - نگاه کنید به Edelman, 2012

^{۱۶}- Social Connectedness Index

کوالو و ریگوبون (۲۰۱۶) از داده‌های قیمتی که از فروشگاه‌های آنلاین استخراج می‌شوند برای ایجاد معیارهایی برای اندازه‌گیری تورم استفاده کردند. این معیارها برای ردیابی آمارهای رسمی مناسب هستند و این مزیت را دارند که می‌توان به دفعات زیاد محاسبه شوند. Goolsbee و Klenow (۲۰۱۸) از مجموعه داده بزرگی از معاملات تجارت الکترونیک برای محاسبه نرخ تورم استفاده کردند. آنها دریافتند که طی دوره ۲۰۱۴-۲۰۱۷، نرخ تورم ۳ درصد کمتر از شاخص رسمی قیمت مصرف کننده^{۱۷} (CPI) بوده است.

یکی دیگر از مجموعه داده‌های بزرگی که اخیراً مورد توجه اقتصاددانان قرار گرفته است داده‌های گوگل ترندز است. این داده‌ها بیانگر میزان جستجوها در موتور جستجوی گوگل در مورد مجموعه ای از کلمات کلیدی در یک منطقه جغرافیایی خاص است. ویژگی کلان داده‌های گوگل ترندز این است که سری‌های زمانی عبارات جستجو شده، نتیجه تجمیع میلیون‌ها پرسش توسط کاربران Google در سراسر جهان است. گوگل ترندز را می‌توان به عنوان معیاری برای اندازه‌گیری احساسات تفسیر کرد، زیرا در یک مقطع زمانی خاص، دغدغه عمومی راجع به یک موضوع خاص را نشان می‌دهد. اولین پژوهش‌های انجام شده بر روی داده‌های گوگل ترندز در مقاله Choi و Varian (۲۰۱۲) منتشر شد؛ نویسندگان دریافتند که بکارگیری ترندها به طور مناسب، دقت پیش‌بینی چندین متغیر اقتصادی را بهبود می‌بخشد. D'Amuri و Marcucci (۲۰۱۷) از بخش جستجوی گوگل راجع به شغل برای پیش‌بینی نرخ بیکاری در ایالات متحده استفاده کردند. نتایج آنها نشان می‌دهد که استفاده از گوگل ترندز دقت را نسبت به پیش‌بینی کنندگان حرفه‌ای بهبود می‌بخشد و به‌ویژه این داده‌ها در نقاط عطفی^{۱۸} که پیش‌بینی در لحظه (real-time) دشوار است، دقیق هستند. در مقاله دیگری، Tran و Castelnuovo (۲۰۱۷) با الهام گرفتن از مقاله بیکر و همکارانش (۲۰۱۶) شاخصی را ایجاد کردند که آن را عدم قطعیت روندهای گوگل^{۱۹} (GTU) می‌نامند که هدف آن ثبت عدم قطعیت اقتصادی و سیاسی^{۲۰} (EPU) با استفاده از داده‌های گوگل ترندز است.

داده‌های دیگر

یکی از کاربردهای جالب داده‌های لرزه‌ای^{۲۱} در اقتصاد توسط Tiozzo Pezzoli و Tosetti (۲۰۲۱) ارائه شده است. آن‌ها از داده‌های لرزه‌ای برای شناسایی ارتعاشات ناشی از فعالیت‌های انسانی، مانند ترافیک هوایی و جاده‌ای و فعالیت‌های تولیدی در میان دیگران استفاده کردند. آنها در مقاله‌شان نشان دادند که شاخصی که می‌سازند با چندین متغیر رسمی فعالیت اقتصادی همبستگی قوی دارد.

منبع دیگری از کلان داده‌ها از تصاویر ماهواره‌ای است که در انواع مطالعات علوم اجتماعی محاسباتی استفاده می‌شود. با این حال، اخیراً اقتصاددانان به پتانسیل داده‌های تصاویر ماهواره‌ای برای تحلیل اقتصادی پی برده‌اند. دونالدسون و استوریگارد

17- Consumer Price Index

18- Turning Points

19- Google Trends Uncertainty

20- Economic and Political Uncertainty

21- seismic data

(۲۰۱۶) و گیبسون و همکاران (۲۰۲۰) در مقالاتشان، مروری بر کاربرد داده‌های ماهواره‌ای در اقتصاد و آغاز سنجش از دور ارائه داده‌اند.

چن و نوردهاوس^{۲۲} (۲۰۱۱) از داده‌های ماهواره‌ای نور شب برای بهبود معیارهای اندازه‌گیری تولید ناخالص داخلی برای کشورهای در حال توسعه استفاده کردند، که مخصوصاً زمانی که آمار رسمی وجود ندارد، قابل استفاده است. این مقاله نشان می‌دهد که درخشندگی، اطلاعات ارزشمندی را ارائه می‌دهد که می‌تواند به بهبودی محاسبات کمک کند. گالیمبرتی^{۲۳} (۲۰۲۰) پژوهش مشابهی را با تمرکز بر توانایی پیش‌بینی معیارهای فعالیت اقتصادی بر اساس داده‌های درخشندگی انجام داد. نتایج نشان می‌دهد که این اقدامات برای بهبود دقت مدل‌های پیش‌بینی ساده^{۲۴}، مفید هستند، اگرچه مدل‌های خاص کشور عملکرد پیش‌بینی بهتری را نسبت به مدل تلفیقی ارائه می‌دهند. هندرسون و همکاران (۲۰۱۱) بحث مفصلی از کاربردهای چراغ‌های شب برای اندازه‌گیری درآمد ملی، به ویژه در مورد اقتصادهای در حال توسعه ارائه داده‌اند. یکی دیگر از کاربردهای داده‌های نور شب توسط استوریگارد (۲۰۱۶) ارائه شده است که ارزیابی می‌کند که آیا فاصله شهرها از یک بندر بر رشد آنها در کشورهای جنوب صحرای آفریقا تأثیر گذاشته است یا خیر. نقش داده‌های ماهواره‌ای در این مورد، ارائه معیاری از فعالیت اقتصادی در سطح شهر است که به طرق دیگر از آمارهای رسمی در دسترس نیست.

زنجیره ارزش کلان‌داده

در حوزه مدیریت بازرگانی، زنجیره ارزش به عنوان یک ابزار پشتیبانی تصمیم برای مدل‌سازی زنجیره فعالیت‌هایی که یک سازمان به منظور ارائه محصول یا خدمات ارزشمند به بازار انجام می‌دهد، بکار برده می‌شود. زنجیره ارزش فعالیت‌های ارزش‌افزای عمومی یک سازمان را دسته‌بندی می‌کند که به آن‌ها اجازه می‌دهد بهینه شوند. یک زنجیره ارزش از یک سری زیرسیستم تشکیل شده است که هر کدام دارای ورودی‌ها، فرآیندهای تبدیل و خروجی هستند. Sviokla و Rayport (۱۹۹۵) یکی از اولین کسانی بودند که در کار خود بر روی زنجیره ارزش مجازی، این استعاره را برای سیستم‌های اطلاعاتی به کار بردند. به عنوان یک ابزار تحلیلی، زنجیره ارزش را می‌توان بر جریان اطلاعات اعمال کرد تا ارزش ایجاد تکنولوژی داده قابل فهم شود. در زنجیره ارزش داده، جریان اطلاعات به عنوان مجموعه‌ای از مراحل مورد نیاز برای تولید ارزش و بینش مفید از داده‌ها توصیف می‌شود به عنوان مثال، کمیسیون اروپا زنجیره ارزش داده را «مرکز اقتصاد دانش آینده» می‌داند که فرصت‌های تحولات دیجیتال را به بخش‌های سنتی‌تر (مانند حمل‌ونقل، خدمات مالی، بهداشت، تولید، خرده‌فروشی) می‌آورد.

22- Chen and Nordhaus

23- Galimberti

24- simple forecasting models

زنجیره ارزش کلان داده، می تواند برای مدل سازی فعالیت های سطح بالا که یک سیستم اطلاعاتی را تشکیل می دهند، استفاده شود. زنجیره ارزش کلان داده فعالیت های کلیدی سطح بالا را که در زیر آمده، شناسایی می کند:

اکتساب داده: اکتساب داده فرآیند جمع آوری، فیلتر کردن و پاکسازی داده ها قبل از قرار دادن آنها در انبار داده یا هر راه حل ذخیره سازی دیگری است که بر اساس آن تحلیل داده ها می تواند انجام شود. اکتساب داده یکی از چالش های بزرگ کلان داده از نظر نیازهای زیرساختی است. زیرساخت مورد نیاز برای پشتیبانی از اکتساب کلان داده ها باید هم در جمع آوری داده ها و هم در اجرای query ها، نتایج با تاخیر کم و قابل پیش بینی ارائه دهد. این زیرساخت باید قادر به انجام حجم معاملات بسیار بالا، اغلب در یک محیط توزیع شده باشد و از ساختارهای داده انعطاف پذیر و پویا پشتیبانی می کند.

تحلیل داده ها: تحلیل داده ها به این می پردازد که داده های خام به دست آمده را قابل استفاده در تصمیم گیری و همچنین قابل استفاده در موضوعات خاص می کند. تحلیل داده ها شامل کاوش، تبدیل و مدل سازی داده ها با هدف برجسته کردن داده های مرتبط، ترکیب و استخراج اطلاعات پنهان مفید با پتانسیل بالا از نقطه نظر تجاری است. حوزه های مرتبط شامل داده کاوی، هوش تجاری و یادگیری ماشین است.

ساماندهی داده^{۲۵}: ساماندهی داده مدیریت فعال داده در طول چرخه عمر آن است تا اطمینان حاصل شود که الزامات کیفی لازم برای استفاده مؤثر از داده برآورده شده است یا نه. فرآیندهای پردازش داده ها را می توان به فعالیت های مختلفی مانند ایجاد محتوا، انتخاب، طبقه بندی، تبدیل^{۲۶}، اعتبار سنجی و حفظ و نگهداری، دسته بندی کرد. پردازش داده ها توسط متصدیان خبره ای انجام می شود که مسئول بهبود دسترسی و کیفیت داده ها هستند. متصدیان داده (همچنین به عنوان متصدیان علمی شناخته می شوند) مسئولیت اطمینان از قابل اعتماد بودن داده، قابل کشف، دسترسی، قابل استفاده مجدد و متناسب با هدف خود بودن داده را بر عهده دارند. یک روند کلیدی برای مدیریت کلان داده از رویکردهای گروه و جمع سپاری استفاده می کند.

ذخیره سازی داده: ذخیره سازی داده ها، تداوم و مدیریت داده ها به روشی مقیاس پذیر است که نیازهای برنامه هایی را که خواستار دسترسی سریع به داده ها هستند برآورده می کند. سیستم های مدیریت دیتابیس رابطه ای^{۲۷} (RDBMS) برای نزدیک به ۴۰ سال راه حل اصلی و تقریباً منحصر به فرد برای پارادایم ذخیره سازی بوده است. با این حال دیتابیس های رابطه ای به دلیل تغییرات طرحواره و عملکرد و تحمل خطا در هنگام افزایش حجم داده ها و پیچیدگی آن، انعطاف پذیری لازم را ندارد و نمی توانند ویژگی های ACID (اتمی^{۲۸}، سازگاری^{۲۹}، جداسازی^{۳۰} و دوام^{۳۱}) که تراکنش های دیتابیس را تضمین می کنند، داشته باشند. در نتیجه دیتابیس های رابطه ای برای

25- Data Curation

26- transformation

27- relational database management systems

28- Atomicity

29- Consistency

30- Isolation

31- Durability

سناریوهای کلان داده نامناسب است. بجای آن، فناوری‌های NoSQL با هدف مقیاس‌پذیری طراحی شده‌اند و طیف وسیعی از راه‌حل‌ها را بر اساس مدل‌های داده‌ی جایگزین ارائه می‌کنند.

استفاده از داده: استفاده از داده، شامل فعالیت‌های تجاری مبتنی بر داده که نیاز به دسترسی به داده‌ها، تحلیل آن و ابزارهای مورد نیاز برای تحلیل داده‌ها را شامل می‌شود. استفاده از داده‌ها در تصمیم‌گیری تجاری می‌تواند رقابت‌پذیری را از طریق کاهش هزینه‌ها، افزایش ارزش افزوده یا هر پارامتر دیگری که می‌تواند بر اساس معیارهای عملکرد موجود اندازه‌گیری شود، افزایش دهد.

اکوسیستم کلان داده

اصطلاح اکوسیستم توسط تانسلی^{۳۲} در سال ۱۹۳۵ برای شناسایی یک واحد اکولوژیکی پایه که شامل محیط و موجودات درون آن است، ابداع شد. در حوزه تجارت، جیمز اف مور^{۳۳} (۱۹۹۳، ۱۹۹۶، ۲۰۰۶) از این استعاره بیولوژیکی استفاده کرد و از این اصطلاح برای توصیف محیط کسب و کار استفاده کرد. مور اکوسیستم کسب و کار را به عنوان «جامعه اقتصادی که توسط بنیادی از سازمان‌ها و افراد در حال تعامل حمایت می‌شود» تعریف کرد. ثابت شده است که یک استراتژی شامل یک شرکت که به تنهایی تلاش می‌کند موفق شود، از نظر ظرفیت آن برای ایجاد محصولات یا خدمات ارزشمند محدود است. بسیار مهم است که کسب و کارها برای بقای خود در یک اکوسیستم تجاری با یکدیگر همکاری کنند (مور ۱۹۹۳؛ گوسین و کندیا ۱۹۹۸). اکوسیستم‌ها به شرکت‌ها اجازه می‌دهند ارزش جدیدی ایجاد کنند که هیچ شرکتی به تنهایی نمی‌تواند به آن دست یابد (ادنر ۲۰۰۶). در یک اکوسیستم کسب و کار سالم، شرکت‌ها می‌توانند با یکدیگر در یک وب تجاری پیچیده کار کنند که در آن به راحتی می‌توانند منابع حیاتی را مبادله و به اشتراک بگذارند. این امر ثابت شده است که اگر یک شرکت به تنهایی برای موفقیت تلاش کند، ظرفیت محدودی برای ایجاد محصول یا خدمات خواهد داشت. در نتیجه همکاری کسب و کارها برای بقای خود در یک اکوسیستم تجاری اهمیت دارد. اکوسیستم‌ها به شرکت‌ها اجازه می‌دهند ارزش جدیدی ایجاد کنند که هیچ شرکتی به تنهایی نمی‌تواند به آن دست یابد. در یک اکوسیستم کسب و کار سالم، شرکت‌ها می‌توانند با یکدیگر در یک وب تجاری پیچیده کار کنند که در آن به راحتی می‌توانند منابع حیاتی را مبادله و به اشتراک بگذارند.

مطالعه اکوسیستم‌های کسب و کار یک حوزه تحقیقاتی فعال است که در آن محققان بسیاری از جنبه‌های استعاری اکوسیستم کسب و کار را بررسی می‌کنند تا جنبه‌هایی مانند اجتماع، همکاری، وابستگی متقابل، تکامل مشترک، عملکردهای اکوسیستمیک و مرزهای محیط‌های تجاری را کشف کنند. کوئینینگ (۲۰۱۲) یک دسته‌بندی ساده از اکوسیستم‌های تجاری را بر اساس درجه کنترل منابع کلیدی و نوع وابستگی متقابل اعضا ارائه می‌دهد. انواع اکوسیستم‌های تجاری شامل سیستم‌های عرضه (مثل نایک)، پلتفرم‌ها (Apple iTunes)، گروه‌های هم‌سرنوشت (مثل Sematech در صنعت نیمه‌رسانا)، و گروه‌های در حال گسترش است.

در اکوسیستم‌های طبیعی، موجودات هوشمند انرژی خود را کنترل می‌کنند. در اکوسیستم‌های تجاری، یک شرکت هوشمند اطلاعات و جریان‌های آن را مدیریت می‌کند. از منظر داده، بکارگیری استعاره اکوسیستم برای توصیف محیط داده‌ای که توسط جامعه‌ای از

³²- Tansley

³³- James F. Moore

سازمان ها و افراد در حال تعامل، پشتیبانی می شود، مفید است. اکوسیستم های کلان داده می توانند به روش های مختلفی در اطراف یک سازمان، پلتفرم های فناورانه، یا درون یا در بین بخش ها شکل بگیرند. اکوسیستم های کلان داده در بسیاری از بخش های صنعتی وجود دارند که در آن حجم وسیعی از داده ها بین بازیگران زنجیره های تامین اطلاعات پیچیده جابجا می شود. بخش های تجاری حاوی اکوسیستم های داده تثبیت شده یا نوظهور عبارتند از مراقبت های بهداشتی، مالی، لجستیک، رسانه، تولید و داروسازی. علاوه بر خود داده ها، اکوسیستم های کلان داده می توانند توسط پلتفرم های مدیریت داده، زیرساخت های داده (مانند پروژه های متن باز مختلف آپاچی) و سرویس های داده نیز پشتیبانی شوند.

یک اکوسیستم داده موفق «صاحبان داده، شرکت های تحلیل داده، متخصصان ماهر داده، ارائه دهندگان خدمات ابری، شرکت هایی از صنایع کاربر، سرمایه گذاران خطرپذیر، کارآفرینان، موسسات تحقیقاتی و دانشگاه ها را گرد هم می آورد». یک اکوسیستم داده موفق، که یکی از ویژگی های برجسته اقتصاد مبتنی بر داده است، می تواند شاهد تعامل یکپارچه این ذینفعان در یک بازار واحد دیجیتال باشد که منجر به فرصت های تجاری، دسترسی آسان تر به دانش و سرمایه می شود.

نمونه موردی: اکوسیستم کلان داده اروپایی

در حالی که تا سال ۲۰۲۰ هیچ اکوسیستم داده منسجمی در سطح اروپا وجود نداشته است، مزایای به اشتراک گذاری و پیوند داده ها در دامنه ها و بخش های مختلف صنعت آشکار شده است. طرح هایی مانند شهرهای هوشمند نشان می دهند که چگونه بخش های مختلف (مانند انرژی و حمل و نقل) می توانند برای به حداکثر رساندن پتانسیل بهینه سازی و بازگشت ارزش با یکدیگر همکاری کنند. بارورسازی متقابل ذینفعان و مجموعه های داده از بخش های مختلف یک عنصر کلیدی برای پیشبرد اقتصاد کلان داده در اروپا است.

اکوسیستم کسب و کار کلان داده اروپایی عامل مهمی برای فهرست بندی تجاری و کالایی سازی خدمات، محصولات و پلتفرم های کلان داده است. یک اکوسیستم کلان داده موفق، شاهد تعامل همه ذینفعان در یک بازار واحد دیجیتال است که منجر به فرصت های تجاری، دسترسی آسان تر به دانش و سرمایه می شود.

یک اکوسیستم داده ای که به خوبی کار می کند، باید سهامداران کلیدی را با منافع آشکار برای همه گرد هم آورد. بازیگران کلیدی در یک اکوسیستم کلان داده، همانطور که در شکل زیر نشان داده شده است، عبارتند از:

- **تأمین کنندگان داده:** شخص یا سازمانی [شرکت های بزرگ و کوچک و متوسط (SME)] که داده ها را از منابع عمومی و خصوصی ایجاد، جمع آوری، ادغام و تبدیل می کند.
- **ارائه دهندگان فناوری:** معمولاً سازمان ها (بزرگ و کوچک و متوسط) به عنوان ارائه دهندگان ابزار، پلتفرم ها، خدمات و دانش برای مدیریت داده ها هستند.
- **کاربران نهایی داده:** شخص یا سازمانی از بخش های مختلف صنعتی (خصوصی و عمومی) که از فناوری و خدمات کلان داده به نفع خود استفاده می کنند.

- **بازار داده:** شخص یا سازمانی که داده های ناشران را میزبانی می کند و به مصرف کنندگان/کاربران نهایی ارائه می دهد.
- **استارت آپ ها و کارآفرینان:** توسعه دهنده فناوری، محصولات و خدمات نوآورانه مبتنی بر داده.
- **محققان و دانشگاهیان:** الگوریتم ها، فناوری ها، روش شناسی، مدل های تجاری و جنبه های اجتماعی جدید مورد نیاز برای پیشبرد کلان داده ها را بررسی می کنند.
- **تنظیم کننده ها** برای حفظ حریم خصوصی داده ها و مسائل حقوقی.
- **نهادهای استانداردساز:** استانداردهای فناوری (چه رسمی و چه بالفعل) را برای ترویج پذیرش جهانی فناوری داده های بزرگ تعریف می کنند.
- **سرمایه گذاران، سرمایه گذاران خطرپذیر و مراکز رشد:** شخص یا سازمانی که منابع و خدماتی را برای توسعه پتانسیل تجاری اکوسیستم ارائه می کند.



شکل ۱- سطوح مختلف اکوسیستم کلان داده

به سمت یک اکوسیستم کلان‌داده

فعال کردن یک اکوسیستم داده مستلزم تعدادی از چالش‌های فنی است که باید با هزینه و پیچیدگی انتشار و استفاده از داده‌ها غلبه کنند. اکوسیستم‌های کنونی با تعدادی از مشکلات مانند کشف داده‌ها، مدیریت، پیوند، همگام‌سازی، توزیع، مدل‌سازی کسب و کار و فروش و بازاریابی مواجه هستند. برای ایجاد اکوسیستم‌های کلان‌داده موثر، باید بر تعدادی از چالش‌های کلیدی اجتماعی و زیست محیطی غلبه کرد. این موارد شامل اما محدود به موارد زیر نیست:

- درک ارزش و سهم فناوری کلان‌داده
- تعیین ارزش داده‌ها
- شناسایی مدل‌های کسب و کاری که از اکوسیستم مبتنی بر داده پشتیبانی کنند
- امکان دسترسی کارآفرینان و سرمایه‌گذاران خطرپذیر برای دسترسی آسان به اکوسیستم
- حفظ حریم خصوصی و امنیت برای همه بازیگران در اکوسیستم
- حقوق مالکیت معنوی در زمینه داده

نتیجه‌گیری

این مباحث نشان می‌دهد که کلان‌داده‌ها چگونه می‌توانند برای پاسخ به سؤالات طولانی مدت و آزمایش اعتبار مفروضات اقتصادی، ارزشمند باشند. یک تصویر، کار با داده‌های اداری است که قبلاً مورد بحث قرار گرفت؛ کار با این داده‌ها پتانسیل بزرگی برای محققان اقتصادی برای کار با داده فراهم می‌کند، اما همچنین این چالش را برجسته می‌کند که دسترسی به این داده‌ها به شکلی فزاینده برای بسیاری از افراد محدودیت دارد. چالش دیگر این واقعیت است که بسیاری از این مجموعه داده‌های جایگزین توسط شرکت‌های خصوصی جمع‌آوری می‌شوند و این شرکت‌ها احتمالاً انگیزه کمی برای به اشتراک گذاشتن داده‌ها با محققان داشته باشند. با این حال، کلان‌داده‌ها نقش عمومی مهمی را ایفا می‌کنند و همین امر نیازمند چارچوبی است که به اشتراک گذاری اطلاعات را تسهیل کند. نمونه‌ای از ارتباط عمومی استفاده از کلان‌داده، تولید شاخص‌های بلادرنگ (real-time) برای کسب و کارها است. در این رابطه، همکاری بین فدرال رزرو و پردازشگر حقوق و دستمزد (ADP) (Cajner et al., 2019) نشان می‌دهد که چگونه کلان‌داده‌های شرکت‌های خصوصی می‌تواند اطلاعات موجود ارائه شده توسط آژانس‌های آماری برای حمایت از سیاست‌های اقتصادی را در لحظه تکمیل کند. این همکاری احتمالاً راه را برای مشارکت‌های گسترده‌تر بین بخش خصوصی و آژانس‌های آماری باز کند. همانطور که بوستیک و همکارانش^{۳۴} (۲۰۱۶) نشان دادند، مدل فعلی تولید داده‌های اقتصادی، در حوزه سازمان‌های دولتی است که سرمایه‌گذاری و اجرای جمع‌آوری داده‌ها، معمولاً در قالب نظرسنجی‌های مصرف‌کننده و کسب و کارها انجام می‌شود. این مدل احتمالاً در آینده تکامل می‌یابد زیرا شرکت‌ها مقادیر فزاینده‌ای از داده‌های اقتصادی را جمع‌آوری می‌کنند که برای تولید آمار رسمی ارزشمند و به احتمال زیاد ارزان‌تر هستند.

34- Bostic et al



همه این موارد نشان می‌دهد که در آینده بسیار نزدیک، بازاری حول کلان‌داده شرکت‌ها ایجاد خواهد شد و این داده‌ها یک زنجیره ارزش مشخصی ارائه می‌کنند؛ همین امر توجه به اکوسیستم کلان‌داده را ضروری می‌کند و مطالعه آنچه در اروپا در حال انجام است می‌تواند یک نمونه مناسب باشد برای آنچه در ایران نیز می‌تواند اجرا شود.

منابع

- V. Novikov, S. (2020) ‘**Data Science and Big Data Technologies role in the Digital Economy**’, *TEM Journal*, pp. 756–762. doi:10.18421/tem92-44.
- Cavanillas, J., Curry, E., Wahlster, W., (2016), **New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe**, Springer.
- Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., Vespe, M., (2023), **Handbook of Computational Social Science for Policy**, Ispara: Springer.