



در جستجوی داده‌ها

مجموعه داده / کلان داده

۳۰۹

علیرضا منسوب بصیری

امروز کافی است سری به شبکه اجتماعی اینستاگرام بزنیم و در نقشه آن، مقصد سفر خود را جست‌وجو کنیم؛ بیش از هزاران عکس از زوایای مختلف در آنجا پیدا می‌شود. تصور کنید ده سال پیش وقتی یک لوح فشرده آموزشی به شاگردان داده می‌شد، هیچ اطلاعاتی نداشتیم که چند تا از بچه‌ها آن را دیده‌اند و از کدام مطالب آن بیشتر استفاده شده است و اگر قرار بود کارایی آن سنجیده شود، صرفاً یک فرم نظرسنجی به ما داده می‌داد؛ اما امروزه از مشاهده شدن یک فیلم آموزشی اطلاعات زیادی دستگیر ما می‌شود. حتی جاهایی که فیلم متوقف شده و دوباره تکرار شده است نیز مشخص می‌گردد و شما به راحتی می‌توانید بفهمید که کدام فیلم‌ها فقط چند ثانیه اولشان دیده شده و بعد رها شده‌اند. این روزها حتی دوربین‌های خیابان‌ها که رانندگان متخلف را جریمه می‌کنند، حاوی اطلاعات زیاد و ذی‌قیمتی هستند که می‌توانند در تصمیم‌گیری‌های مختلف به ما کمک کنند. تصمیم‌گیری‌هایی که به برکت سرعت زیاد تولید داده، حجم زیاد اطلاعات و تنوع آن است. این اطلاعات که به «بیگ دیتا»

مقدمه

ده سال پیش که مجله رشد مدرسه فردا چاپ می‌شد، دوربین دیجیتال من ۱۲۸ مگابایت حافظه داشت و من در یک سفر، بالغ بر دویست عکس با آن می‌انداختم. هنوز فراموش نکرده‌ام دورانی را که در یک سفر، بیست و چهار یا نهایتاً سی و شش عکس می‌انداختم. امروزه به راحتی در یک سفر بیش از پانصد عکس می‌اندازیم؛ فیلم هم می‌گیریم. تازه، روی عکس‌ها مختصات محل هم ذخیره می‌شود و بسته به نوع دوربین، حتی می‌توان سرعت شاتر، میزان دیافراگم و اطلاعات و داده‌های متنوع دیگر را نیز همراه عکس ذخیره کرد که بعدها مورد رجوع و استفاده قرار بگیرند. سی سال پیش وقتی می‌خواستیم مقصد سفر انتخاب کنیم، مجلات گردشگری و آلبوم سفرهای دوستانمان بود که می‌شد در آن‌ها عکسی از مقاصد گردشگری دید و تصمیم گرفت. البته ده سال پیش وضع بهتری داشتیم؛ با یک جست‌وجو در اینترنت به تعداد محدودی وبلاگ و وبگاه دسترسی پیدا می‌کردیم که اغلب عکس‌های یکدیگر را کپی کرده بودند. اما

یا «کلان داده» معروف است، در تمام زمینه‌ها از علم گرفته تا ورزش، تا اقتصاد، تا بهداشت و درمان، تفکر ما و تعامل ما با دنیا را تغییر داده است. این روزها از بیگ دیتا یا کلان داده در ادبیات حوزه فناوری زیاد استفاده می‌شود و دیتا ساینتیست از مشاغل پرطرفدار کشورهای پیشرفته شده است و احتمالاً به زودی در ایران نیز جزء مشاغل می‌شود که مخاطبان زیادی خواهد داشت و افراد نخبه را به سمت خود جذب خواهد کرد.

کلان داده

تعریف دقیق و کامل از کلان داده مجموعه داده‌هایی است که اندازه آن‌ها فراتر از حدی است که با نرم‌افزارها و روش‌های معمول بتوان آن‌ها را در یک زمان قابل قبول، دریافت، ذخیره، مدیریت و پردازش کرد. کلان داده حاصل اطلاعاتی است که روزبه‌روز عموماً به‌خاطر تولید داده توسط تجهیزات و ابزارهای مختلف دیجیتال در حال افزایش است. گوشی‌های موبایل، حسگرهای محیطی، لاگ نرم‌افزارهای مختلف، دوربین‌ها، میکروفون‌ها، ایستگاه‌های هواشناسی، سامانه‌های امواج رادیویی، مبادلات شبکه‌های اجتماعی برخط (آن‌لاین)، متون و اسناد اینترنتی، داده‌های نجوم، اطلاعات پزشکی و سلامت بیماران، اطلاعات سامانه‌های خرید از فروشگاه‌ها، پژوهش‌های زمین‌شناسی و غیره نمونه‌هایی از داده‌ها در مقیاس کلان هستند؛ مقیاسی که امروزه از گیگابایت و ترابایت به پتابایت و اگزابایت و زتابایت در حال حرکت است.

طراحی فعالیت‌های پژوهش‌محوری که مبتنی بر حجم زیادی از اطلاعات متنوع و گسترده است و جمع زیادی از دانش‌آموزان در جمع‌آوری آن نقش داشته‌اند، شروع خوبی برای کار با داده‌های بزرگ است

- کلان داده‌ها دارای سه ویژگی مهم هستند:
- **حجم داده:** حجم داده‌های درون‌سازمان و خارج آن به مدد پدیده اینترنت، دستگاه‌های الکترونیکی و موبایل‌ها، زیرساخت‌های شبکه و سایر منابع هر ساله رشد نمایی دارد و پیش‌بینی شده است که تا سال ۲۰۲۰ ده زتابایت داده در جهان خواهیم داشت.
- **سرعت تولید:** داده‌ها از طریق برنامه‌های کاربردی و سنسورهای بسیار زیادی که در محیط وجود دارند، با سرعتی بسیار زیاد و به‌صورت بی‌وقفه تولید می‌شوند که اغلب باید در لحظه پردازش و ذخیره شوند.
- **تنوع:** انواع منابع داده و تنوع در نوع داده

بسیار زیاد است؛ در نتیجه، ساختارهای داده‌ای بسیار زیادی وجود دارد و بیشتر حجم داده دنیا هم بی‌ساختار و بسیار متنوع است. بخشی از داده‌ها امروزه در بانک‌های اطلاعاتی، بخشی در صفحات وب و بخشی دیگر در فایل‌ها با قالب‌های متفاوت ذخیره شده‌اند که عمل پردازش آن‌ها را پیچیده می‌کند. تفاوت‌های اصلی کلان‌داده و داده‌های سنتی در جدول مقابل بیان شده است.

معیار	داده‌های سنتی	کلان داده
اندازه	گیگابایت تا ترابایت	پتابایت تا اگزابایت
معماری	متمرکز	توزیع شده
ساختار	دارای ساختار	بی‌ساختار یا نیم‌ساختار
مدل داده	مدل داده ثابت	بدون شمای مشخص
ارتباط داخلی	ارتباطات پیچیده بین رکوردها	فاقد ارتباطات داخلی پیچیده

کلان داده و تحول آموزش علم

شاگردان ما قصد ادامه تحصیل در هر رشته‌ای که داشته باشند، تجزیه و تحلیل کلان‌داده‌ها برایشان مهارت مهمی خواهد بود، اما آیا با روش‌های متداول آموزش علوم این مهارت در آن‌ها رشد پیدا می‌کند و یا تقویت می‌شود؟ در آموزش علوم متداول، معمولاً یک پدیده در کلاس به مشاهده گذاشته می‌شود و حجم و تنوع داده‌هایی که در کلاس معمولی تولید می‌شوند به اندازه مشاهده‌هایی است که فقط داده تولید می‌کنند و نه مجموعه‌ای از داده‌ها. بنابراین، دانش‌آموزان فرصت نمی‌یابند تا تجربه‌هایی را کسب کنند که مناسب نسل خودشان باشد. مهارت‌های فرایندی آموزش علوم که بیشتر معلمان در کلاس درس روی آن‌ها تأکید دارند و در طراحی آموزشی کلاس خود آن‌ها را لحاظ می‌کنند، به تنهایی نمی‌تواند شاگردانی تربیت کند که قرار است در آینده با حجم عظیمی داده مواجه شوند که ساختار ندارند، غیرمتمرکزند و در ظاهر به هم ارتباط ندارند. مهارت «اندازه‌گیری و داده‌ها» در ریاضیات و «تجزیه و تحلیل و تفسیر داده‌ها» در علوم تجربی و مهندسی از جمله مهارت‌های استانداردهای آموزش علوم برای نسل آتی هستند که در چند سال اخیر مطرح شده‌اند. در راستای مهارت «تجزیه و تحلیل و تفسیر داده‌ها» دانش‌آموز باید:

- یاد بگیرد که از نمودار و جدول برای نشان دادن اطلاعات استفاده کند.
- بتواند در میان دسته‌ای از داده‌ها الگویابی کند و رابطه بین داده‌ها را پیدا کند.
- بتواند فرضیه‌های خود را متناسب با داده‌ها بازنگری کند.
- یاد بگیرد که براساس داده‌ها و شواهد مستند تصمیم‌گیری کند.

راهکارهای آمادگی برای کلان داده

برای تبدیل دانش‌آموز به یک کاربر کلان‌داده ابتدا باید توجه داشت که تا پیش از این او با چه داده‌هایی سروکار داشته است. دانش‌آموزی که هنوز حتی با مجموعه‌ای کوچک از داده‌ها کار نکرده و همواره در معرض یک نوع داده بوده است، نمی‌تواند در یک گام بزرگ وارد کلان داده شود. او می‌بایست اول یاد بگیرد

