بزرگ اندیشی

گذر از داده‌های کوچک به داده‌های بزرگ

کیم کاستنز[1]**، روت کرامنزل**[2] **وآیرین بیکر**[3]
ترجمه و تلخیص: **محمدامین اسپروز**

داده‌هـا شـالودهٔ علم‌انـد. هر بینـش و حقیقـت علمی که درکتاب‌ها موجود اسـت ریشه در مسـتنداتی دارد که بر پایهٔ داده‌ها بنا شـده‌اند. اسـتانداردهای علمی نسل آتی[4] «تجزیه و تحلیل و تفسیر داده‌ها» را به عنوان یکی از هشت روش علمی و مهندسـی معرفی می‌کند و اسـتانداردهای مشـترک‌المرکز ایالتی در ریاضیات[5] «اندازه‌گیری و داده‌ها» را به عنوان یکی از حوزه‌هـای ریاضیات ـ که باید در تمامی مقاطع تحصیلی رواج یابد ـ می‌شناسد.

دانشـمندانی چـون *گالیله*، *فـارادی* و *کوری* معمـولاً روی داده‌هایی کار می‌کردند که خودشان شخصاً و یا به کمک گروه کوچکـی آن‌ها را جمع‌آوری کرده بودند. امروزه، اکتشـافات علمی بیشـتر از طریق بررسی مجموعهٔ پیچیده‌ای از داده‌ها محقق می‌شـود که جمع‌آوری آن‌ها توسـط یک دانشمند به تنهایی ممکن نیست.

این دسـت از داده‌ها ـ برای نمونه، داده‌های ژنتیکی DNA که در مرکز ملی اطلاعات فناوری زیستی[6] موجود است ـ ممکن است با همکاری تعداد زیادی آزمایشگاه جمع‌آوری شده باشد و یا مانند داده‌های زمین‌شناسی ارت‌اسکوپ[7] با هماهنگی ابزاری مرکزی جمع‌آوری شـده باشـد. در هر صـورت، داده‌ها به‌طور مرکزی بایگانی می‌شوند و به رایگان و به صورت برخط (آنلاین) در دسترس قرار می‌گیرند.

داده‌هایـی که اکنون در اختیار معلمـان علوم و دانش‌آموزان اسـت، از لحاظ کیفی و کمی بـا داده‌هایی که پیش‌تر تنها در مؤسسـه‌های پژوهشـی موجود بود، برابری می‌کند. مجموعهٔ بـزرگ داده‌ها در سـایر زمینه‌ها، از معامـلات املاک گرفته تا بازاریابـی، آموزش، حقوق کیفری و ورزش، نیز می‌تواند کاربرد داشته باشد.

شاگردانتان به دنبال هر رشته‌ای که باشند، تجزیه و تحلیل داده‌های بزرگ برایشان توانایی مهمی خواهد بود

شـاگردانتان به دنبال هر رشته‌ای که باشند، تجزیه و تحلیل داده‌های بزرگ برایشان توانایی مهمی خواهد بود.

تمرکز این مقاله بر معلم‌هایی اسـت که پیش از ایـن تجربـهٔ کار روی داده‌هـای کوچـک و جمع‌آوری شده توسط دانش‌آموزان را داشته‌اند و حال می‌خواهند بـا داده‌های آنلاین بزرگ‌تر، که دانشمندان و مهندسان جمع‌آوری کرده‌اند، کار کنند. ما در اینجا به چالش‌ها، راهکارهای آموزشی و منابع مناسب درسی خواهیم پرداخت.

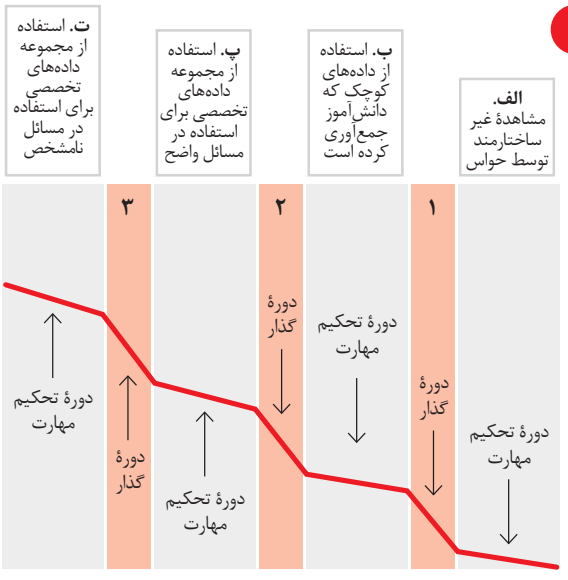
چالش‌ها و تغییرات

در ابتدا، کودکان با استفاده از حواسشان جهان را بدون ساختار مشاهده می‌کنند (بخش الف). آن‌ها می‌آموزند که پدیده‌هایی را که به‌طور مسـتقیم تجربه می‌کنند، پیش‌بینی کنند و یا حکم مربوط به آن‌ها را تعمیم‌دهند؛ برای مثال، «آب در سراشـیبی جاری می‌شود». سپس، به‌عنوان دانش‌آموز یاد می‌گیرند که با داده‌های کوچکی که خود جمع‌آوری کرده‌اند، کار کنند (بخش ب). اکثر تحقیق‌های کلاسی متعارف ـ از جمله تحقیق‌هایی که

در استانداردهای ملی آموزش علوم[8] آمده است ـ در این بخش جای می‌گیرند. نمونهٔ این دسته، کاشت گیاهان بر لبهٔ پنجرهٔ کلاس و بررسی شرایط مختلف آبیاری است.

بعدها، دانش‌آموزان با مجموعه داده‌های بزرگ‌تری کار می‌کنند که خود آن را جمع‌آوری نکرده‌اند. در بیشتر مواقع، این داده‌ها به‌صورت برخط (آنلاین) به‌دست می‌آیند. در ابتدا آن‌ها روی مسئله‌هایی واضح مانند موقعیت‌یابی مرز گسل‌ها با استفاده از پایگاه داده‌های زلزله‌ها و آتش‌فشان‌ها کار می‌کنند (بخش پ) و در آخر، یاد می‌گیرند که با استفاده از مجموعه‌داده‌هایی بزرگ و پیچیده روی مسائلی غیر ساختارمند کار کنند. برای مثال، این مسئله که آیا استفاده از شکستگی هیدرولیکی برای جامعهٔ ما مفید است و یا کدام رسیدگی پزشکی برای یکی از بستگان مسن مناسب است (بخش ت). شکل ۱ مسیری را نشان می‌دهد که طی آن دانش‌آموز با کسب مهارت در یک بازه و پس از دوره‌ای برای گذار، به مرحلهٔ بعدی یادگیری وارد می‌شود.

۱



مسیر ارتقا به یک کاربر داده‌های حرفه‌ای

نویسندگان این مقاله معتقدند که مسیر تبدیل کودک به یک کاربر داده‌های حرفه‌ای به دوره‌های گذار چالش برانگیزی نیاز دارد. تمرکز این مقاله برگذار از یک کاربر داده‌های کوچک به یک کاربر حرفه‌ای داده‌های بزرگ است.

این مقاله دربارهٔ گذار از مجموعه داده‌های کوچکی که دانش‌آموز جمع‌آوری کرده به مجموعه‌داده‌های بزرگی است که به‌صورت تخصصی جمع‌آوری شده و در شکل یک به‌عنوان گذار ۲ آمده‌اند. در این گذار، تغییرات و چالش‌های بسیاری وجود دارد.

داده‌هایی که دانش‌آموزان جمع‌آوری می‌کنند معمولاً از چند صد نقطه داده تجاوز نمی‌کند؛ حال آنکه داده‌هایی که متخصصان جمع‌آوری می‌کنند، با واحد گیگابایت اندازه‌گیری می‌شوند. مجموعه داده‌های دانشمندان معمولاً بازهٔ زمانی گسترده‌تری را پوشش می‌دهد و میزان نمونه‌برداری، مراکز نمونه‌برداری و جامعهٔ آماری بزرگ‌تری را در برمی‌گیرد. همچنین، شرایط و گونه‌های ترکیبی و متفاوتی از داده‌ها را شامل می‌شود. دانش‌آموزان عموماً داده‌هایشان را در جدول‌ها و یا نمودارهای ساده به نمایش می‌گذارند اما برای نمایش مجموعه داده‌های بزرگی که دانشمندان جمع‌آوری کرده‌اند، به ابزار آماری و یا تجسمی پیچیده‌تری نیاز است.

به علاوه، زمانی که دانش‌آموزان خود به جمع‌آوری داده‌ها می‌پردازند، فهم عمیق‌تری از روند تولید داده و محدودیت‌هایش پیدا می‌کنند. آن‌ها ممکن است نسبت به شرایط و یا محیط‌هایی که داده از آن‌ها استخراج شده است، حسی شخصی پیدا کنند و این می‌تواند بر تحلیل آن‌ها از این داده‌ها تأثیر بگذارد. تمامی

این موارد وقتی داده‌ها را شخص دیگری جمع‌آوری کرده باشد، صدق نمی‌کنند.

تغییرات و چالش‌های گذار به داده‌های بزرگ		
قبل		بعد
داده‌ها را دانش‌آموز جمع‌آوری کرده است.		داده‌ها را دانشمندان، مهندسان و فناوران به‌صورت تخصصی جمع‌آوری کرده‌اند.
مجموعهٔ کوچکی از داده‌ها		مجموعهٔ بزرگی از داده‌ها
مشاهدات تجربی و ساده (مانند اندازه‌گیری بارش باران توسط باران‌سنج)		شاخص‌های مرکب (مانند اندازه‌گیری میزان بارش براساس داده‌های ماهواره‌ای)
دانش مستقیم نسبت به موضوع و محیطی که داده‌ها از آن مستخرج شده توسط حواس شخصی		درکی پراکنده و انتزاعی از محیط توسط فراداده (مانند تاریخ/ زمان، طول/ عرض جغرافیایی، موقعیت در نقشه)
مشارکت در طراحی جمع‌آوری داده‌ها باعث ایجاد آگاهی مستقیم می‌شود که گستردگی را قربانی عمق می‌کند.		دانشی دست دوم دربارهٔ اینکه داده‌ها چگونه جمع‌آوری شده‌اند.
از برخی مشکلات احتمالی داده‌ها از طریق تجربهٔ شخصی آگاه می‌شود (مانند از بین رفتن داده‌ها،خطای اندازه‌گیری، ناکارآمدی ابزار، خطاهای عملیاتی).		باید از طریق مؤلفه‌های خود داده‌ها و یا فراداده مشکلات احتمالی را شناسایی کند.
ابزاری ساده و قابل فهم برای اندازه‌گیری: دفتر و مداد، کاغذ گراف، ماشین حساب		ابزار نمایش‌داده‌ها و یا نرم‌افزارهای آماری قدرتمند و درعین حال مبهم

راهکارهای آموزشی

با اینکه گذار از مجموعه داده‌های کوچک به مجموعه داده‌های بزرگی که به‌صورت تخصصی جمع‌آوری شده چالش برانگیز است اما برای دانش‌آموزان دبیرستان امری ممکن است. در ادامه، ما چهار راهکار آموزشی را که در کلاس امتحان شده‌اند آورده‌ایم. این راهکارها به دانش‌آموزان در اولین برخوردشان با داده‌های بزرگ و تخصصی کمک می‌کنند.

راهکار شمارهٔ ۱: پازل داده‌ها: شما می‌توانید ازاین راهکار در تمامی درس‌ها استفاده کنید و برای اجرای آن نیازی به فناوری نیست. به دنبال فعالیت‌ها و پرسش‌هایی باشید که دانش‌آموزان برای پاسخ‌گویی به آن‌ها باید داده‌های تصویری (برای مثال، نمودارها و نقشه‌ها) را که دانشمندان جمع‌آوری کرده‌اند، تحلیل کنند. در یک پازل داده‌ها، که به خوبی طراحی شده است، تنها برشی از داده‌ها انتخاب می‌شود؛ زیرا این برش می‌تواند تجسم واضحی از ساختار و یا روندی مهم را ارائه دهد و برای دانش‌آموزان حجم عظیمی از یادگیری ـ نسبت به تلاش ـ دارد. برای مثال، دانش‌آموزان می‌توانند تجسم روند تکامل را در نوک فنچ‌های جزایر گالاپاگوس[۹] بررسی کنند. گروهی از این پرندگان در اثر یک خشکسالی عظیم مردند و گروهی دیگر از آن جان سالم به در بردند. به منظور ارزیابی درک دانش‌آموزان، با سؤالاتی داده‌محور و ساده شروع کنید. سعی کنید برخی از سؤالاتتان، به جای اینکه به خواندن اعداد و ارقام از روی نقشه یا نمودار محدود شود، دانش‌آموز را به تحلیل روندی که داده‌ها نشان می‌دهند ملزم کند. در مثالی که آورده شد، سؤالی که می‌تواند دانش‌آموزان را به رمزگشایی و توصیف داده‌ها باز دارد ممکن است این باشد: «انحراف متوسط اندازهٔ نوک را در دو گونه فنچ با هم مقایسه کنید.» سؤالی که به تفسیر نیاز داشته باشد می‌تواند این باشد: «براساس دانشتان از علوم زیستی و با استفاده از داده‌های موجود، فکر می‌کنید انحراف متوسط اندازهٔ نوک در این دو گونه فنچ چه دلیلی می‌تواند داشته باشد؟»

زمانی که دانش‌آموزان خود به جمع‌آوری داده‌ها می‌پردازند، فهم عمیق‌تری از روند تولید داده و محدودیت‌هایش پیدا می‌کنند. آن‌ها ممکن است نسبت به شرایط و یا محیط‌هایی که داده از آن‌ها استخراج شده است، حسی شخصی پیدا کنند و این می‌تواند بر تحلیل آن‌ها از این داده‌ها تأثیر بگذارد. تمامی این موارد وقتی داده‌ها را شخص دیگری جمع‌آوری کرده باشد، صدق نمی‌کنند

یکی از معلم‌هایی که در کلاسش از مثال فنچ‌ها استفاده کرده بود می‌گوید: « به طرز عجیبی چالش برانگیز بود. بعضی از دانش‌آموزان در همان ابتدا مسئله را فهمیدند و توانستند با استفاده از داده‌ها به سؤالات تحلیلی پاسخ دهند. برخی دیگر کمی به مشکل برخوردند و نیاز داشتند دوباره به داده‌ها رجوع کنند. من برای راهنمایی از آن‌ها چند سؤال کردم تا بتوانند روابط داده‌ها را درک کنند. دفعهٔ بعد، بیشتر خود را کنار خواهم کشید تا دانش‌آموزان زمان بیشتری به تحلیل سؤالات و داده‌ها بپردازند.»

راهکار شمارهٔ ۲: مجموعه داده‌های تو در تو: دانش‌آموزان ابتدا مجموعهٔ کوچکی از داده‌ها را خود جمع‌آوری و تفسیر می‌کنند. سپس، به تفسیر مجموعهٔ بزرگ‌تری از داده‌ها

می‌پردازند. برای مثال، گروه‌هایی از دانش‌آموزان ابتدا از طریق جمع‌آوری داده‌ها در نزدیکی مدرسه، دربارهٔ خلیج دهانه‌ای رود هادسون[10] تجربهٔ شخصی پیدا می‌کنند. سپس با استفاده از داده‌های جمع‌آوری شده توسط گروه‌های دیگر مدرسه به بالا و پایین رودخانه و از طریق داده‌های جمع‌آوری شدهٔ تخصصی به‌وسیلهٔ حسگری دائمی در رودخانه، تحقیق خود را در طول زمان گسترش می‌دهند.

برای ارزیابی، سؤالاتی را که با داده‌های جمع‌آوری شده توسط دانش‌آموزان قابل پاسخ‌گویی است با سؤالاتی که به مجموعهٔ بزرگ‌تری از داده‌ها نیاز دارند، ترکیب کنید. در مثالی که آورده شد، یک سؤال مربوط به داده‌های کوچک می‌تواند این باشد که « در نقطهٔ نمونه‌برداری ما، میزان شوری آب در هنگام افزایش جزر و مد چگونه تغییر کرد؟» و سؤال مربوط به مجموعهٔ بزرگ داده‌ها: «به عقیدهٔ شما سه عاملی که در میزان شوری آب تأثیر می‌گذارند چیست؟ پاسختان را با استفاده از داده‌ها تقویت کنید.»

پیش‌بینی کردن باعث می‌شود دانش‌آموزان از نتیجه سهمی داشته باشند. به علاوه، آن‌ها را به مشاهدهٔ دقیق‌تر تشویق و در بررسی پایگاه داده‌ها راهنمایی می‌کند

راهکار شمارهٔ ۳: پیش‌بینی، مشاهده، توضیح: این تسلسل آموزشی بیشتر برای فعالیت‌های عملی کاربرد دارد و می‌تواند به دانش‌آموزان در بررسی مجموعهٔ بزرگی از داده‌ها کمک کند. با توجه به درک دانش‌آموزان از فرایندهای طبیعی، آن‌ها می‌توانند چگونگی مجموعهٔ داده‌ها در مجموعه شرایطی مشخص را پیش‌بینی کنند. برای مثال، دانش‌آموزان پیش‌بینی می‌کنند که اگر ستاره‌ای، سیاره‌ای داشته باشد که به دور آن بگردد، ناظر زمینی باید بتواند کم شدن حجم نور به دلیل قرارگیری سیاره میان زمین و ستاره را مشاهده کند. پیش‌بینی کردن باعث می‌شود دانش‌آموزان از نتیجه سهمی داشته باشند. به علاوه، آن‌ها را به مشاهدهٔ دقیق‌تر تشویق و در بررسی پایگاه داده‌ها راهنمایی می‌کند.

به منظور ارزیابی، فهم دانش‌آموزان از شباهت‌ها و تفاوت‌های میان پیش‌بینی و داده‌های مشاهده شده را واکاوی کنید. وقتی دانشمندان از الگوها استفاده می‌کنند، این شباهت‌ها و تفاوت‌ها به آن‌ها نشان می‌دهد که کدام الگو موفق بوده است و کدام الگو به اصلاح نیاز دارد. برای مثال، این سؤال‌ها را می‌توانید مطرح کنید:«داده‌ها از چه جهاتی با آنچه انتظار داشتید مشابه بودند؟» « از چه جهاتی متفاوت بودند؟» «فکر می‌کنید این شباهت‌ها و تفاوت‌ها چه دلیلی دارد؟»

راهکار شمارهٔ ۴: فرضیه‌سازی: در مواردی که دانش‌آموزان ممکن است اطلاعات کافی برای پیش‌بینی نظام‌مند نداشته باشند، شما می‌توانید چندین فرضیه را در اختیار آن‌ها بگذارید. آنان داده‌های موجود را در راستای یکی از فرضیه‌ها بررسی و دسته‌بندی می‌کنند. در یک نمونه، فرضیه‌ها در قالب طرح‌هایی از شکل زمین هستند که ممکن است در پایگاه‌داده‌های زمین‌شناسی/ عمق‌سنجی وجود داشته باشد. تحقیقات نشان می‌دهد که در اختیار گذاشتن این سطح از احتمالات باعث می‌شود که دانش‌آموزان داده‌ها را روشمندتر و مفیدتر بررسی کنند.

برای ارزیابی، کافی نیست که دانش‌آموزان فرضیهٔ درست را انتخاب کنند بلکه باید بتوانند انتخابشان را با استناد به داده‌ها تحلیل کنند. در نمونه‌ای که نام برده شد، یک سؤال خوب این است که «برای هر یک از فرضیه‌هایی که رد کردید، با استفاده از داده‌ها نشان دهید که آن شکل زمین نمی‌تواند شکل زمینی باشد که مورد تحقیق قرار گرفته است.»

منابع داده‌ها و درس‌ها

با وجود راهکارها، شما هنوز برای موضوعاتی که می‌خواهید تدریس کنید به برنامهٔ درسی و داده‌ها نیاز دارید. خوشبختانه، داده‌ها بیش از پیش به‌صورت برخط (آنلاین) در دسترس قرار گرفته‌اند، ابزارهای دسترسی به داده‌ها ساده‌تر شده‌اند و برنامه‌های درسی داده ـ محور افزایش یافته‌اند. در شکل ۳، به معرفی منابع معتبری که در کلاس امتحان شده‌است می‌پردازیم. دانش‌آموزان با مراجعه به این منابع و با استفاده از داده‌هایی که دانشمندان جمع‌آوری کرده‌اند، می‌توانند روی موضوعات مختلفی چون آب و هوا، زمین‌شناسی، تکامل و ژنتیک کار کنند.

نتیجه‌گیری

دانش‌آموزان با راهنمایی شما و استفاده از داده‌های برخط و توانایی تازه‌شان در استفاده از داده‌های بزرگ، می‌توانند به جای پذیرش کورکورانهٔ نظریه‌های علمی ، مستنداتی برای برخی از آن‌ها پیدا کنند. آن‌ها می‌توانند به بررسی فضاها و شرایطی بپردازند که خودشان ممکن است هیچ‌گاه به آنجا نروند؛ مثلاً اعماق اقیانوس. آن‌ها می‌توانند پدیده‌هایی را مورد بررسی قرار دهند که جهان شمول‌اند و در طول سال‌ها ادامه

داشـــته‌اند. همچنیـــن می‌توانند از جمله افرادی باشـــند که از همان داده‌ها اســـتفاده کرده‌اند یا خودشان دست به اکتشاف بزنند. تجهیز دانش‌آموزان با داده‌هایی پیچیده برای حرفه‌های رو به گسترش، از مکانیک گرفته تا پزشکی، مفید است. برای مشاغل مرتبط با علوم و مهندسی، نیز داشتن مهارت در این امر ضروری است.

*** پی‌نوشت**

1. Kim Kastens
2. Ruth Krumhansl
3. lrene Baker
4. The Next Generation Science Standards
5. Common Core State Standards, Mathematics
6. National Center For Biotechnology Information
7. EarthScope: http://www.earthscope.org/
8. National Science Education Standards
9. Galapagos Islands

منابعی برای فعالیت‌های درسی با استفاده از داده‌هایی که به‌صورت تخصصی جمع‌آوری شده‌اند.

منابع	نمونه‌ای از فعالیت‌ها
The Center for Innovation in Engineering and Science Education, Stevens Institute of Technology (*http://ciese. org/realtimeproj.html)*	é آلودگی هوا: راه‌حل چیست؟ é سفر بر جریان یک خلیج é موج‌ها é خروش سونامی
Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of wisconsin Madison (*https://cimss.ssec.wisc.edu/satmet/index.html*)	é بادهای ماهواره‌ای é پیش‌بینی آب و هوا é تصاویر ماهواره‌ای é نظارت بر محیط زیست‌جهانی
NOAA (*http://datainthe classroom.noaa.gov/Dataln The Classroom/*)	é بررسی ارتفاع از سطح دریا با داده‌های واقعی é درک اسیدی شدن اقیانوس‌ها é نتیجه‌گیری: نقشه‌های آب و هوا
EDC (*http://oceantracks.org*) (*http://oceansofdata.org/our-work/edcearth-science*)	é بررسی مهاجرت‌های موجودات دریایی با تغییر فصل é الگوی جریان‌های اقیانوس é مقایسهٔ دماهای منطقه‌ای
Earth Exploration Toolbook (*http://serc, carleton.edu/eet/index.html*)	é بررسی حرکات گسل‌ها از طریق داده‌های GPS é تاریخ وضعیت آب و هوایی از طریق رسوب‌های کف دریا é وجود آب
Evolution and the Nature of Science Institutes (*www.indiana.edu/~ensiweb/lessons/p.tut.db.html*)	é بررسی مسائل مربوط به تکامل با استفاده از پایگاه داده‌های آنلاین مولکولی é بررسی ردپاهای لاتولی é تاریخ‌نگاری رسوبی
Northwest Association for Biomedical Research (*http://bit.ly/1xXoAWw*)	é چگونه بیوانفورماتیک در آزمایش‌های ژنتیکی کاربرد می‌یابد؟ é چگونه از بیوانفورماتیک در پژوهش‌های ژنتیکی استفاده می‌شود؟
Bioinformatics Activity Bank (*http://bit.ly/19jv67E*)	é چه چیز باعث سرگیجه می‌شود؟ é رنگ چشم: آیا آبی، واقعاً آبی است؟ é تکامل خرس‌ها