



بزرگ داده (big Data) عبارتی است که مدتهاست درباره آن می‌شنویم اما هنوز ابهاماتی در رابطه با ماهیت حقیقی این بزرگ داده‌ها وجود دارند. در حقیقت، بزرگ داده‌ها نیروی محرکه اصلی در ظهور تکنولوژی‌های جدید از جمله هوش مصنوعی، علم دیتا و اینترنت اشیا (IoT) در عصر دگرگونی دیجیتالی هستند.

با درنظر گرفتن این تعاریف، ما نیز تصمیم به انتشار تعریفی جامع و ابتدایی از ماهیت بزرگ داده‌ها در سال 2017 کردیم. همانند گذشته، این آموزش نیز به صورت کاملا جامع و با قصد توصیف مفاهیم اصولی بدون توجه به دانش قبلی خوانندگان درباره بزرگ داده‌ها تهیه و تنظیم شده است.

همه چیز با آغاز عصر داده‌ها و تحول عظیم در حجم داده‌های تولیدی آغاز گردید. این گسترش حجم داده‌ها به علت توسعه کامپیوترها، اینترنت و تکنولوژی که قادر است اطلاعات را از دنیای واقعی و فیزیکی دریافت و آنها را به داده‌های دیجیتال تبدیل کند، به وجود آمد.

امروزه، هر بار که آنلاین می‌شویم، هر بار که اسمارت‌فون‌های مجهز به جی‌پی‌اس را با خودمان حمل می‌کنیم، هر بار که از طریق رسانه‌های اجتماعی و اپلیکیشن‌های چت با دوستانمان ارتباط برقرار می‌کنیم و حتی زمانی که خرید می‌کنیم، در اصل دیتای جدید تولید می‌کنیم. بهتر است بگوییم هر عملی که شامل انجام فعالیت‌های دیجیتالی شود، ردپای دیجیتالی کاربر را در آنجا بر جای خواهد گذاشت.

موضوع مهمتر مقدار داده‌های تولید شده توسط ماشین‌ها است که با سرعت بسیار زیاد در حال افزایش است. زمانی که تجهیزات هوشمند منازل با یکدیگر یا با سرور مرکزی منزل ارتباط برقرار می‌کنند، داده‌ها تولید و به اشتراک گذاشته می‌شود. در سراسر دنیا، هر روزه تعداد بیشتری از ماشین آلات صنعتی کارخانه‌ها به سنسورهایی که داده‌ها را مخابره می‌کنند، مجهز می‌شوند. به زودی نیز خودروهای بدون راننده به خیابان‌ها می‌آیند و به سرعت نقشه‌ی چهاربعدی فضای اطراف خودرو را به سرورهای مرکزی انتقال خواهند داد.

وظیفه بزرگ داده‌ها چیست؟

اطلاعات روزافزون به دست آمده از سنسورها، تصاویر، متن، اطلاعات صوتی و ویدیویی، همگی تشکیل‌دهنده‌ی بزرگ داده‌هایی هستند که امروزه می‌توانیم از آنها به روش‌هایی که چند سال پیش غیرممکن بود، استفاده کنیم. در حال حاضر، پروژه‌های مرتبط با تحلیل بزرگ داده‌ها در بخش‌های زیر در حال کمک‌رسانی هستند.

درمان بیماری‌ها و جلوگیری از سرطان

داروهای داده محور (Data-Driven) شامل تحلیل حجم گسترده‌ای از سوابق دارویی جهت یافتن الگوهای با قابلیت کمک به دانشمندان جهت شناسایی بیماری‌ها در مراحل ابتدایی و تولید داروهای جدید می‌شود.

غذاسازی به گرسنگان

به لطف داده‌ها، کشاورزی در حال متحول شدن است و به زودی با استفاده از بزرگ داده‌ها می‌توان تولید محصولات را به حداکثر و حجم آلاینده‌های ورودی به محیط زیست را به حداقل رساند و استفاده از ماشین‌ها و ابزارآلات را بهینه‌سازی کرد.

جستجوی سیاره‌های دوردست

ناسا با آنالیز میلیون‌ها داده از آنها برای مدل‌سازی تمامی امکانات ممکن جهت فرود مریخ‌نوردهای خود بر روی این سیاره و برنامه‌ریزی جهت انجام ماموریت‌های آینده استفاده می‌کند.

پیش‌بینی و مقابله با بلایای طبیعی و انسانی

با تحلیل داده‌های به دست آمده از سنسورها می‌توان محل احتمالی زمین‌لرزه‌های بعدی را پیش‌بینی و با آنالیز الگوهای رفتاری انسان‌ها می‌توان به سرخ‌هایی که قادر است به

سازمان‌ها جهت نجات جان افراد و کمک‌رسانی بیشتر به بازماندگان کمک کنند، دست پیدا کرد. همچنین با به‌کارگیری تکنولوژی تحلیل بزرگ‌داده‌ها می‌توان جهت حفاظت از جان مهاجران در مناطق جنگی سراسر دنیا استفاده کرد.

جلوگیری از وقوع جرم

نیروهای پلیس بر اساس استراتژی‌های برگرفته شده از داده‌های قبلی و مجموعه داده‌های عمومی، قادر خواهند بود نیروهای خود را موثرتر وارد عملیات کرده و در مواقع ضروری با بازدارندگی بیشتری در مقابل مجرمین وارد عمل شوند.

به لطف وجود بزرگ‌داده‌ها، امکان خرید آنلاین، انتخاب بهترین وسیله و مکان جهت سفر، انتخاب بهترین زمان برای رزرو بلیط هواپیما و تصمیم‌گیری جهت انتخاب فیلم مناسب برای تماشا در سینما و در کل زندگی به شکل ساده و راحت‌تر امکان‌پذیر شده است.

قاعده اصلی بزرگ‌داده‌ها به این شکل است که هرچه بیشتر درمورد یک مسئله یا یک موقعیت اطلاعات به‌دست بیاورید، بینش درونی شما نسبت به آن بیشتر و پیش‌بینی‌های شما درباره اتفاقات بعدی با اطمینان بیشتری حاصل می‌شود. با مقایسه داده‌ها، روابطی که در گذشته غیرقابل تشخیص بود کم‌کم آشکار می‌شوند و در ادامه، این روابط درک و آگاهی ما از تصمیمات آینده را برایمان امکان‌پذیر خواهند کرد.

عموماً، این نتایج از طریق مدل‌سازی بر اساس داده‌های جمع‌آوری شده، انجام مراحل شبیه‌سازی، محاسبه مقدار داده‌ها در هر مرحله و بررسی تاثیر آنها بر روی نتایج کسب شده، به‌دست می‌آیند. امروزه این مراحل به‌صورت اتوماتیک و از طریق تکنولوژی‌های تحلیلی پیشرفته انجام می‌گیرد که قادرند شبیه‌سازی را میلیون‌ها بار به صورت مکرر انجام دهند و تمامی متغیرها را تا زمانی که به یک الگوی قابل فهم دست پیدا کنند، جابه‌جا کنند و از این طریق مسئله مورد نظر را در مدت کوتاهی حل کنند.

هرروزه، داده‌ها به شکل سازمان‌دهی نشده وارد زندگی‌مان می‌شوند. منظور از سازماندهی نشده این است که نمی‌توان آنها را به راحتی در جدول‌های دسته‌بندی شده قرار داد. بخش اعظم این داده‌ها به شکل تصاویر و ویدیوهای هستند که از تصاویر به‌دست آمده از ماهواره‌ها گرفته تا عکس‌های آپلود شده در فیس‌بوک و توییتر را تشکیل می‌دهند. ایمیل‌ها، پیام‌های متنی اپلیکیشن‌های پیام‌رسان و مکاتبات تلفنی ضبط شده را نیز بخش دیگر این داده‌ها را تشکیل می‌دهند. برای درک تمامی این داده‌ها، پروژه‌های تحلیلی بزرگ‌داده‌ها با استفاده از به‌روزترین تکنولوژی‌های تحلیل اطلاعات از جمله هوش مصنوعی و یادگیری ماشینی انجام می‌شوند. با آموزش کامپیوترها به منظور درک این داده‌ها از طریق تشخیص تصاویر و پردازش به زبان طبیعی، کامپیوترها می‌توانند الگوها را با سرعت و اطمینان بیشتری نسبت به انسان‌ها تشخیص دهند.

روند پیش‌گرفته شده در چند سال اخیر موجب پیشرفت‌های چشمگیر در این زمینه و ارائه ابزارهای تحلیل بزرگ‌داده‌ها به عنوان پلتفرمی جهت خدمت‌رسانی شده است. سازمان‌ها و شرکت‌ها فضای سرور، سیستم‌های نرم‌افزاری و پردازشگرها را از ارائه دهندگان سرویس‌های کلود اجاره می‌کنند. تمامی مراحل تحلیل داده‌ها از طریق سیستم‌های شرکت ارائه‌دهنده این سرویس‌ها انجام می‌شود و مشتریان تنها هزینه انجام آن را پرداخت می‌کنند. این نوع همکاری‌ها، تحلیل و تبدیل بزرگ‌داده‌ها را برای تمامی سازمان‌ها ممکن ساخته و از صرف هزینه‌های هنگفت جهت خرید سخت‌افزار، نرم‌افزار و استخدام کارکنان تخصصی جدید جلوگیری می‌کند.

نگرانی‌ها درباره بزرگ‌داده‌ها

امروزه، بزرگ‌داده‌ها بینش و فرصت‌های بی‌سابقه‌ای را در اختیارمان قرار می‌دهند اما در این میان سوالات و نگرانی‌هایی وجود دارند که باید به آنها اشاره کرد.

حریم خصوصی داده‌ها

بزرگ‌داده‌هایی که امروزه تولید می‌کنیم حاوی حجم زیادی از اطلاعات زندگی شخصی افراد که بایستی در حریم خصوصی محافظت شوند، هستند. در حال حاضر هم اطلاعات خصوصی افراد را فاش می‌کنیم و هم از فضای راحتی که برنامه‌ها و سرویس‌های مرتبط با بزرگ‌داده‌ها فراهم کرده‌اند استفاده می‌کنیم. حقیقتاً چه کسی اجازه دسترسی به این اطلاعات را دارد؟

امنیت داده‌ها

حتی اگر با به‌کارگیری اطلاعاتمان برای مقاصد به‌خصوص مشکلی نداشته باشیم، آیا حفظ این اطلاعات توسط افراد تضمین می‌شود؟ آیا قوانین و مقررات مرتبط با حفظ اطلاعات خصوصی برای این مقیاس از داده‌ها نیز کاربرد دارند؟

تبعیض اطلاعاتی

زمانی که همه چیز مشخص شد، آیا تبعیض قائل شدن برای برخی افراد بر اساس اطلاعاتی که از زندگی شخصی آنها به دست آورده‌ایم قابل قبول خواهد بود؟ در حال حاضر، بانک‌ها بر اساس رده‌بندی اعتبار افراد، تصمیم می‌گیرند که به چه افرادی وام اعطا کنند و سازمان‌های بیمه نیز شدیداً به داده‌ها اتکا کرده‌اند. باید انتظار داشته باشیم که اطلاعات با جزئیات گسترده‌تری ارزیابی و تحلیل شوند و از اینکه در اختیار داشتن اطلاعات فردی موجب سخت‌تر شدن معیشت افرادی که از منابع و دسترسی به اطلاعات محدودتری برخوردارند نمی‌شود، اطمینان حاصل کنیم.

مواجهه با چنین چالش‌هایی، بخشی از مشکلات بزرگ‌داده‌ها محسوب می‌شوند. بدون شک اختلافات گسترده‌ای بین دانشمندان و متفکرین دانشگاه‌ها وجود دارد و افرادی که قصد استفاده از بزرگ‌داده‌ها را دارند، باید از مخاطرات و ایرادات این کار توسط این افراد باخبر شوند. در غیر این صورت، کسب‌وکارها با صدمات جدی روبرو شده و در ادامه بحران‌های مالی و جریمه‌های سنگین را به همراه خواهد داشت.

اوایل که افراد درباره بزرگ‌داده‌ها صحبت می‌کردند، بسیاری آن را تکنولوژی زودگذری می‌دانستند که قرار است تنها برای مدتی کوتاه مورد توجه قرار بگیرد و زمانی که تکنولوژی جدیدتری ظهور کند، بزرگ‌داده‌ها نیز فراموش می‌شوند. اما تا به امروز چنین اتفاقی نیفتاده است و در حقیقت، با ظهور تکنولوژی‌های جدید، می‌بینیم که نیروی محرکه اصلی تمامی آنها بزرگ‌داده‌ها هستند. در آینده نزدیک، مقدار داده‌های در اختیار ما گسترش پیدا می‌کنند و تکنولوژی‌های تحلیل داده‌ها نیز پیشرفته‌تر می‌شوند. پس اگر بزرگ‌داده‌ها امروزه قادر به انجام چنین کارهای بزرگی هستند، تصور کنید در آینده قادر به انجام چه کارهای بزرگتری خواهند بود.

منبع: [آی تی ایران](#)