

به نام خدا

شهاب الدین لاهوتی
آزمایشگاه پایگاه داده ها
استاد یاریان

Big Data

Big Data چیست؟

بیگ دیتا یک واژه باب روز (Buzzwords) برای فرایندهایی شامل جمع آوری، استخراج و تحلیل داده‌هایی پیچیده است که با ابزارهای معمول نمی‌توان از آنها برای دستیابی به اهداف استفاده کرد.

این تعریف ساده بیگ دیتا همه ماجرا را بیان نمی‌کند. در واقع اصطلاح بیگ دیتا فقط به داده‌ها اشاره می‌کند در حالی که بخش بزرگی از این فرایند بر عهده علم داده (Data science)، استخراج داده (data mining)، تحلیل داده (Data analysis)، یادگیری ماشین (Machine learning) و در ابعاد بزرگتر، هوش مصنوعی است.

کاربرد های Big Data

آموزش

بیگ‌دیتا در هر مرحله از آموزش یعنی از کودکی تا آموزش عالی یا هر دوره آموزشی دیگر می‌تواند کاربرد متنوعی داشته باشد.

برای مثال؛ می‌توان انتخاب رشته و کاریابی را برای دانشجویان آسانتر کرد، به جای برنامه‌های همسان آموزشی برای هر دانشجو ترتیب یادگیری دروس متفاوتی با بهره‌وری بالاتر پیشنهاد داد یا با شناسایی نقاط ضعف دوره‌های آموزشی از انصراف و رد شدن دانشجویان جلوگیری کرد.

رسانه و تلویزیون

سریال‌های تلویزیونی که در سال‌های گذشته منتشر شده‌اند داستان‌هایی به مراتب جذاب‌تر دارند. به علاوه پیدا کردن یک سریال که با روحیات و علایق هر فرد مطابقت داشته باشد کار چندان سختی نیست.

یک نمونه مشهور استفاده از بیگ دیتا در سریال‌های تلویزیونی پروژه **کمپانی نتفلیکس** است. این شرکت با کمک بیگ دیتا علایق مردم را شناسایی می‌کند تا در وهله اول سریال‌های بهتری را به مخاطبان پیشنهاد دهد و سپس از این داده‌ها در ساخت سریال‌های جدید خود استفاده کند.

ویژگی های Big Data

۱. حجم زیاد (Volume)

۲. نرخ تولید بالا (Velocity)

۳. تغییر پذیری و تنوع زیاد (Variety)

۴. صحت داده‌ها (Veracity)

۵. ارزش (value)

انواع تحلیل Big Data

۱. تحلیل توصیفی (Descriptive)

۲. تحلیل تشخیصی (Diagnostic)

۳. تحلیل پیش‌بینی (Predictive)

۴. تجویزی (Prescriptive)

آینده Big Data

متخصصین علوم داده می‌گویند هر انسانی که در حال حاضر در شهری پیشرفته زندگی می‌کند و به لوازم دیجیتال دسترسی دارد در هر ثانیه حدود ۵/۱ مگابایت داده تولید می‌کند. این رقم هر روز بزرگتر می‌شود و این داده‌ها با سرعت بیشتری ذخیره می‌شود.

بیگ دیتا در معنای دقیقش چیز عجیبی نیست. در واقع ما انسان‌ها می‌دانستیم که با داشتن داده‌های بیشتر می‌توان ارتباط پدیده‌ها را کشف کرد، اما امکانات و بستری برای این کار نداشتیم.

به لطف اینترنت و ظهور سایر فناوری‌ها این بستر فراهم شد و در آینده‌ای نزدیک با گسترش بیشتر اینترنت اشیاء سرعت و حجم تولید داده افزایش چشمگیری خواهد داشت.

در این که کسب‌وکارها ناچارند به سمت فرایندهای داده‌محور بروند شکی نیست اما آن‌هایی موفق خواهند شد که هرچه زودتر وارد این عرصه شوند.