

به نام خدا

دانشجو: محمد مهدی رضائی

نام استاد: میثاق یاران

موضوع: big data

کلان داده (big data) چیست:

مجموعه‌ای بسیار بزرگ از داده‌ها است که از منابع متعددی جمع‌آوری می‌شود. این داده‌ها می‌توانند نتایج ارزیابی عملکرد یک سازمان یا تعاملات مخاطبان آن در شبکه‌های اجتماعی باشد. کلان داده‌ها می‌توانند سرنوشت یک سازمان را به کلی زیرورو کنند. تجزیه و تحلیل این داده‌ها در یک سازمان می‌تواند بینشی ایجاد کند که منجر به تصمیم‌گیری‌های بهتر و کشف مسیرهای استراتژیک تجاری شود

انواع کلان داده به سه نوع داده‌های ساختاریافته، بدون ساختار و نیمه ساختاریافته است که در ادامه به آن می‌پردازیم.

1- داده‌های بدون ساختار:

تمام داده‌های سازمان نیافته ما، داده‌های بدون ساختار هستند. تقریباً ۸۰٪ از داده‌های سراسر جهان ساختار ندارند. تقریباً هر کاری که شما با کامپیوتر انجام می‌دهید، داده‌های بدون ساختار تولید می‌کند. این داده‌ها، ماهیت پیچیده‌ای دارند، فضای بیشتری را اشغال می‌کنند و بی‌نظمی آن مدیریت درک آن را دشوار می‌کند. سخت‌ترین قسمت تحلیل داده‌های بدون ساختار این است که به یک برنامه یاد بدهیم که اطلاعات به دست آمده را درک کند. برای این کار باید اطلاعات را برای برنامه، به فرم‌هایی از داده‌های ساختاریافته ترجمه کرد که کارآسانی نیست، معمولاً برای این کار از روش‌هایی مانند تجزیه متن، پردازش طبیعی زبان و... استفاده می‌کنند.

2- داده‌های ساختاریافته:

تصور کنید صدها صفحه داده، در ستون‌ها و ردیف‌هایی مرتب شده‌اند. برای تمام عنوان‌ها توضیح وجود دارد، متغیرها را به راحتی می‌شود تشخیص داد و اعداد قابل درک و ملموس هستند. معلوم است که کار با این داده‌ها آسان است و برای برنامه‌ریزی

به راحتی می توان داده ها را مرتب و جمع آوری کرد. برای مثال یک پایگاه داده حقوق و دستمزد کارمندان، داده هایی مثل اطلاعات شناسایی کارکنان، نرخ دستمزدشان، ساعت کار، نحوه پرداخت اضافه کاری و غیره را به ما می دهد. الگوی داده های ساختاریافته، برای هر برنامه ای که از آن استفاده می کنیم، هر یک از این موارد را تعریف خواهد کرد. پس این برنامه برای کشف معنای واقعی هر کدام از این موارد، مجبور نیست به جستجوی داده بپردازد، یعنی می تواند مستقیماً به جمع آوری و پردازش آن بپردازد.

2- داده های نیمه ساختاریافته:

ترکیبی از داده های ساختاریافته و بدون ساختار هستند. برای این داده ها الگوهای معنادار و جدول های مخصوص طراحی نشده است. اما برای داده ها برچسب ها و نشانه هایی برای معنا کردن وجود دارند که داده ها را برای ضبط و طبقه بندی و ساخت پرونده در مجموعه های داده آماده می کند که ذخیره سازی آن را نسبت به داده های ساختار نیافته آسان تر می کند.

کاربردهای کلان داده چیست:

کلان داده در حوزه های زندگی بشر کاربر دارد. کاربرد داده ها فقط برای شرکت ها، سازمان های بزرگ و دولت ها نیست. به طور مثال **کلان داده در صنعت سلامت** به شناسایی بیماری در مراحل اولیه کمک می کند، در **بخش های دولتی** که روزانه با حجم زیادی از داده ها روبرو می شوند. آنها باید سوابق و پایگاه های اطلاعاتی مختلفی را در مورد شهروندان، رشد جامعه در ابعاد مختلف، منابع انرژی، نقشه برداری های جغرافیایی و... را ثبت و پایش کنند، کلان داده ها در هواشناسی که پروژه تحقیقاتی **IBM Deep Thunder**، با تجزیه و تحلیل پیچیده کلان داده ها، امکان پیش بینی دقیق آب و هوا را فراهم کرده است. از این پیش بینی ها برای آمادگی و ایمنی در برابر بلایای طبیعی و پیش بینی احتمال خرابی خطوط برق در شهر توکیو استفاده شده است. و صنعت حمل و نقل و داده ها و بانکداری و..

