

داده یا دیتا (Data) چیست؟

داده یا دیتا (Data) یا دیتا به هر مجموعه‌ای از حقایق، آمار یا اطلاعاتی اطلاق می‌شود که ذخیره می‌شود و برای تجزیه و تحلیل یا به عنوان مرجع استفاده می‌شود. داده می‌تواند شکل‌های مختلفی داشته باشد و معمولاً به صورت ساختاریافته یا بدون ساختار سازماندهی می‌شود. داده‌ها مبنایی را برای درک الگوها، روندها و روابط فراهم می‌کنند و ما را قادر می‌کند تا بینش‌های ارزشمندی به دست آوریم و تصمیم‌های آگاهانه بگیریم.

تجزیه و تحلیل داده چیست؟

تجزیه و تحلیل داده‌ها (Data Analysis) فرایند سیستمیک بازرسی، پاک‌سازی، تبدیل و مدل‌سازی داده‌ها برای کشف الگوهای معنادار، استخراج بینش‌های ارزشمند و تصمیم‌گیری آگاهانه است. این شامل به کارگیری تکنیک‌های مختلف آماری و محاسباتی برای کشف رابطه‌ها، شناسایی روندها و کشف الگوهای پنهان در مجموعه داده‌هاست. تجزیه و تحلیل داده‌ها شامل فعالیت‌هایی مانند پیش‌پردازش داده‌ها، آمار توصیفی، آمار استنباطی، تجسم داده‌ها و مدل‌سازی پیش‌بینی می‌شود. با انجام تجزیه و تحلیل داده‌ها، سازمان‌ها و افراد می‌توانند بینش‌های ارزشمندی به دست آورند فرایندها را بهینه کنند، ناهنجاری‌ها را شناسایی کنند، و تصمیم‌گیری مبتنی بر شواهد را در حوزه‌های مختلف هدایت کنند.

انواع داده یا دیتا (Data)

داده‌ها از جنبه‌های مختلفی به انواع متفاوتی تقسیم می‌شوند. مانند تصویری گفتاری و..... است.

داده‌های متنی

داده‌های متنی اطلاعات نوشته شده یا تایپ شده، از پیام‌های کوتاه تا اسناد طولانی، را شامل است. این نوع دیتا شامل ایمیل‌ها، پست‌های رسانه‌های اجتماعی، مقالات، کتاب‌ها و موارد دیگر است. تحلیل داده‌های متنی شامل تکنیک‌های پردازش زبان طبیعی (NLP) برای استخراج معنا، تحلیل احساسات، مدل‌سازی موضوع، بازیابی اطلاعات و طبقه‌بندی متن است. درک داده‌های متنی برای کارهایی مانند تجزیه و تحلیل احساسات مشتری، توصیه‌ی محتوا، تشخیص اسپم و ترجمه زبان ضروری است.

داده‌های تصویر

داده‌های تصویری اطلاعات بصری گرفته‌شده از طریق عکس‌ها، گرافیک‌ها یا اسکن‌ها را نشان می‌دهد. در حوزه‌های مختلفی از جمله بینایی کامپیوتر، تصویربرداری پزشکی و تجزیه و تحلیل تصویرهای ماهواره‌ای استفاده می‌شود. تجزیه و تحلیل داده‌های تصویر شامل وظایفی مانند طبقه‌بندی تصویر، تشخیص اشیاء، بخش‌بندی تصویر (Image Segmentation) و تولید تصویر است. مدل‌های یادگیری عمیق، مانند شبکه‌های عصبی کانولوشن (CNN)، با دستیابی به عملکرد قابل توجهی در کارهایی مانند تشخیص تصویر، شرح تصویرها و انتقال سبک، تجزیه و تحلیل تصویر را متحول کرده‌اند.

داده‌های صوتی یا گفتاری

داده‌های صوتی یا گفتاری (Speech) شامل ضبط صدا، از جمله گفتار، موسیقی یا سایر سیگنال‌های صوتی است. تجزیه و تحلیل داده‌های صوتی شامل تشخیص گفتار (Speech Recognition)، شناسایی گوینده، طبقه‌بندی ژانر موسیقی و تشخیص احساسات است. تکنیک‌هایی مانند تبدیل فوری و مدل‌های مبتنی بر یادگیری عمیق مانند شبکه‌های عصبی بازگشتی (RNN) و شبکه‌های عصبی کانولوشنی (CNN) برای پردازش و استخراج اطلاعات معنی‌دار از سیگنال‌های صوتی استفاده می‌شوند.

داده‌های سری زمانی

داده‌های سری زمانی (Time Series) مجموعه‌ای از نقاط داده است که در فواصل زمانی منظم در طول زمان جمع‌آوری می‌شوند. در زمینه‌هایی مانند امور مالی، پیش‌بینی آب‌وهوا، تجزیه و تحلیل بازار سهام و تجزیه و تحلیل داده‌های حسگر رایج است. تجزیه و تحلیل داده‌های سری زمانی شامل تشخیص روندها، الگوها و ناهنجاری‌ها است. تکنیک‌های پیش‌بینی سری‌های زمانی، مانند میانگین متحرک یکپارچه اتورگرسیون (ARIMA)، هموارسازی نمایی، و شبکه‌های عصبی بازگشتی (RNN)، معمولاً برای پیش‌بینی بر اساس الگوهای تاریخی استفاده می‌شوند.