

مقدمه مؤلفین

خداوند متعال را سپاسگزاریم که توفیق تدوین این مجموعه را برایمان فراهم نمود. کتابی که پیش رو دارید به شرح یکی از مفاهیم نوین داده کاوی به نام "داده های کلان" یا "Big Data" پرداخته است. کتاب حاضر می تواند به عنوان راهنما برای دانشجویان مهندسی صنایع، مهندسی فن آوری اطلاعات، مهندسی کامپیوتر، مدیریت صنعتی، مدیریت بازرگانی، مدیریت اجرایی و ... در مقطع کارشناسی و تحصیلات تکمیلی استفاده شود. به علاوه این کتاب مورد کاربرد عموم محققین، مدیران، کارشناسان و تحلیلگران حوزه های مختلف کسب و کار است. تحلیلگران کسب و کار افرادی هستند که به معرفی، مدیریت و نظارت بر فرایندهای کسب و کار سازمان می پردازند. این افراد عموماً متخصصان فن آوری اطلاعات و یا برنامه ریزانی هستند که فرایندهای کسب و کار را درون سیستم های مدیریت فرایند کسب و کار به اجرا در می آورند.

کتاب حاضر در هشت بخش تنظیم شده است. بخش اول به مفاهیم مقدماتی داده های کلان، بخش دوم به معرفی فن آوری های وابسته و بخش سوم به شرح تولید و دریافت داده های کلان اختصاص دارد. در بخش چهارم ذخیره داده های کلان و در بخش پنجم تحلیل داده های کلان ارائه شده است. در بخش ششم به برنامه های داده های کلان، در بخش هفتم به مسائل حل نشده و چشم انداز و در بخش آخر به بیان واژه نامه پرداختیم.

در خاتمه لازم می دانیم از تمامی بزرگوارانی که به نحوی ما را در تهیه و تألیف این کتاب یاری نمودند، کمال تشکر و قدردانی را به عمل آوریم. امید است خوانندگان عزیز با ارائه نظرات و پیشنهادات سازنده خود ما را مورد لطف و عنایت خویش قرار دهند.

عباس سقایی
مریم نیکوپور

abbas.saghaei@gmail.com
nikoopour.mn@gmail.com

پیشگفتار

تغییرات سریع تکنولوژی در دهه اخیر و افزایش چشمگیر نقش اطلاعات، نحوه مدیریت و رهبری را دگرگون ساخته و اداره سازمان ها به روش های سنتی تقریباً ناشدنی است. محیط امروز، سازمانها و مدیران را با چالش های جدیدی روبرو می سازد و هر روز پویاتر از قبل می شود. موسساتی که خواهان افزایش سهم خود در بازار و کسب منافع بیشتری هستند باید خود را با تغییرات محیط تطبیق دهند.

سازمانها جهت اتخاذ تصمیمات راهبردی تجاری به داده ها متکی اند. اما "داده های کلان" چالش های خاص خود را دارد و می بایست پیش از آنکه اثر واقعی خود را بر جای بگذارند، بررسی شوند. داده های کلان به مجموعه داده هایی اطلاق می شود که مدیریت، کنترل و پردازش آنها فراتر از توانایی ابزارهای نرم افزاری در یک زمان مورد انتظار است. یکی از مهم ترین مسائل مرتبط با داده های کلان، مشکل بودن کار با آن ها به وسیله پایگاه های داده ای رابطه ای و بسته های نرم افزاری تصویرنگاری داده ها و نرم افزارهای آماری رومیزی است. چرا که این داده ها برای پردازش شدن در یک زمان معقول به نرم افزارهای به شدت موازی شده با قابلیت اجرا روی ده ها، صدها یا هزاران سرور نیاز دارند.

امروزه به دلیل تولید و جمع آوری داده ها از مجموعه بزرگی از تجهیزات و ابزارهای مختلف (مانند گوشی های موبایل، حسگرهای محیطی، دوربین ها، میکروفون ها، دستگاه های تشخیص امواج رادیویی، شبکه های حسگر بی سیم و غیره) حجم داده های ذخیره شده در پایگاه ها و انبارها با سرعت خیره کننده ای در حال افزایش است. نکته جالب توجه در این زمینه آن است که ۹۰ درصد داده هایی که اکنون در اختیار ما قرار دارد، تنها در ۲ سال اخیر تولید شده اند! سازمان های بزرگ هیچ گاه مانند اکنون داده های عظیمی نداشته اند و یا حداقل به فکر جمع آوری و استفاده از آن نبوده اند. به همین دلیل، تعجبی ندارد که ابزارهای ما در زمینه مدیریت چنین داده های کلانی پیشرفت چندانی نکرده باشد!