داده کاوی

مفاهیم، مدل‌ها، روش‌ها و الگوریتم‌ها

ویراست سوم

مؤلف

محمد کانتارزیک

مترجم

ایوب ترکیان

نیاز دانش

فهرست مطالب

عنوان **شماره صفحه**

فصل ۱

مفاهیم داده‌کاوی

۱.۱ مقدمه

علوم و مهندسی نوین مبتنی بر استفاده از مدل‌های قاعده- اول (فرمول‌پایه) برای شرح سیستم‌های فیزیکی، بیولوژیکی، و اجتماعی هستند. این‌گونه رویکرد با یک مدل علمی پایه، نظیر قوانین نیوتن حرکت یا معادلات ماکسول در الکترومغناطیس، شروع شده، و آنگاه کاربردهای مختلف را در مهندسی مکانیک و مهندسی برق بنا می‌نهد. در این رویکرد، داده‌های تجربی برای صحه‌گذاری مدل‌های فرمول‌پایه نهشته شده و تخمین بعضی از پارامترهایی که مشکل بوده یا در بعضی مواقع برای اندازه‌گیری مستقیم محال است، استفاده می‌شوند. با این حال، در بسیاری از زمینه‌ها، فرمول‌پایه نهشته معلوم نبوده، یا سیستم‌های تحت بررسی برای ساختاردهی ریاضی بیش از حد پیچیده هستند. با رشد استفاده از کامپیوترها، مقدار زیادی داده توسط این‌گونه سیستم‌ها در حال تولید است. در غیاب مدل‌های فرمول‌پایه، این‌گونه داده‌های به سادگی موجود را می‌توان برای استخراج مدل‌ها با تخمین روابط مفید بین متغیرهای یک سیستم (یعنی، وابستگی‌های ورودی-خروجی نامعلوم)، استفاده کرد. *بدین‌طریق، در حال حاضر یک شیفت پارادایمی از مدل‌سازی و تحلیل‌های مبتنی بر قواعد فرمول‌پایه به توسعه مدل‌ها و تحلیل‌های متناظر از داده‌ها در حال وقوع است.*

به تدریج به این واقعیت عادت کرده‌ایم که حجم بالایی از داده‌های اشغال‌کننده کامپیوترها، شبکه‌ها، و زندگی ما وجود دارد. دوایر دولتی، نهادهای علمی، و بیزینس‌ها، منابع عظیمی را برای