

خلاصه

یادگیری ماشین پتانسیل زیادی برای بهبود محصولات، فرایندها و تحقیقات دارد. اما رایانه‌ها معمولاً پیش‌بینی‌های خود را توضیح نمی‌دهند که مانعی برای پذیرش یادگیری ماشین است. این کتاب درباره تفسیر مدل‌های یادگیری ماشین و تصمیمات آن‌هاست.

پس از بررسی مفاهیم تفسیرپذیری، با مدل‌های ساده و قابل تفسیری مانند درخت تصمیم، قوانین تصمیم‌گیری و رگرسیون خطی آشنا خواهید شد. تمرکز کتاب بر روی روش‌های مدل-آگنوستیک برای تفسیر مدل‌های جعبه سیاه مانند اهمیت ویژگی و اثرات محلی انباشته شده و توضیح پیش‌بینی‌های فردی با مقادیر Shapley و LIME است.

همه روش‌های تفسیر به طور عمیق توضیح داده شده و به صورت انتقادی مورد بحث قرار می‌گیرند. چگونه زیر روپوش کار می‌کنند؟ نقاط قوت و ضعف آن‌ها در چیست؟ چگونه می‌توان خروجی‌های آن‌ها را تفسیر کرد؟ این کتاب شما را قادر می‌سازد تا روش تفسیری را که برای پروژه یادگیری ماشین شما مناسب‌تر است، انتخاب و به درستی اعمال کنید. خواندن این کتاب برای یادگیران یادگیری ماشین، دانشمندان داده، آماردانان و هر کسی که علاقه‌مند به تفسیرپذیر ساختن مدل‌های یادگیری ماشین است، توصیه می‌شود.

درباره من: نام من Christoph Molnar است، من یک آماردان و یک متخصص یادگیری ماشین هستم. هدف من این است که یادگیری ماشین را قابل تفسیر کنم.

من را در توییتر دنبال کنید! @ChristophMolnar

جلد توسط @YvonneDoinel

همچنین کتاب دوم من ذهنیت‌های مدل‌سازی را بررسی کنید.

این کتاب تحت مجوز Creative Commons Attribution-NonCommercial-ShareAlike 4.0 بین‌المللی مجوز دارد.