به نام خدا

متن‌کاوی به کمک
یادگیری ماشین

نویسندگان: چارو سی. آگراوال

مترجم:
دکتر مهدی اسماعیلی

متن‌کاوی به کمک یادگیری ماشین

مترجم: دکتر مهدی اسماعیلی
ناشر: انتشارات آتی‌نگر
ناشر همکار: وینا
صفحه‌آرایی و طراحی جلد: همتا بیداریان
تیراژ: ۵۰۰ نسخه
چاپ اول: ۱۳۹۸
قیمت: ۹۹۰,۰۰۰ ریال
شابک: ۹-۵۹-۶۱۰۲-۶۲۲-۹۷۸

ISBN: 978-622-6102-59-9

نشانی دفتر فروش: خیابان جمالزاده جنوبی، روبه‌روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱
تلفن: ۸-۶۶۵۶۵۳۳۶ **نمابر:** ۶۶۵۶۵۳۳۷
www.ati-negar.com ∗ info@ati-negar.com

سرشناسه: آگراوال، چارو سی، ۱۹۷۰– م. Aggarwal، Charu C
متن‌کاوی به کمک یادگیری ماشین/ نویسنده چارو سی. آگراوال، مترجم: مهدی اسماعیلی
تهران: آتی‌نگر، وینا ۱۳۹۸
۵۹۲ ص.: مصور، جدول، نمودار.
ISBN: 978-622-6102-59-9
وضعیت فهرست نویسی: فیپا.
یادداشت: عنوان اصلی کتاب: Machine Learning for Text, 2018
یادداشت: کتابنامه.
یادداشت: نمایه.
موضوع: فراگیری ماشینی– Machine learning– متن‌پردازی– Text processing (Computer science)
موضوع: هوش مصنوعی– Artificial intelligence– داده‌کاوی– Data mining
شناسه افزوده: اسماعیلی، مهدی، ۱۳۵۰– ، مترجم
شناسه‌افزوده: بیداریان، همتا، ۱۳۶۱– ، گرافیست
رده‌بندی کنگره: Q۳۲۵/۵
رده‌بندی دیویی: ۰۰۶/۳۱
شماره کتابشناسی ملی: ۵۸۱۵۴۴۳

فهرست مطالب

ناظر روی تو صاحب نظری نیست که نیست
بوی گیسوی تو در هیچ سری نیست که نیست

پیشگفتار مترجم

در سال‌های اخیر به دلیل افزایش متون در محیط‌هایی نظیر وِب، رسانه‌های اجتماعی و دیگر پلت‌فُرم‌ها، متن‌کاوی از اهمیت ویژه‌ای برخوردار شده است. در تحلیل متن، تکنیک‌هایی از حوزه‌های دیگر نظیر بازیابی اطلاعات، یادگیری ماشین و پردازش زبان طبیعی دیده می‌شود، و برای هر یک از آن‌ها نیز کتاب‌های زیادی نوشته شده است. تمرکز این کتاب بر روی الگوریتم‌های یادگیری ماشین برای اسناد متنی است؛ هر چند در برخی از فصل‌ها رنگ روش‌های حوزه‌های دیگر پررنگتر شده است. به زعم مترجم، کتاب حاضر یکی از ارزنده‌ترین کتاب‌هایی است که در این حوزه به رشته تحریر درآمده است. به همین دلیل از میان کتاب‌های موجود در این حوزه، به انتخاب و ترجمه آن همت گماشتیم.

مطالب کتاب در چهارده فصل گردآوری و تهیه شده است که می‌توان آن را در سه بخش گروه‌بندی کرد. بخش اول یعنی فصل‌های اول تا هشتم کتاب، شامل الگوریتم‌ها و مدل‌های پایه برای تحلیل متن است. تجزیه ماتریس، خوشه‌بندی و رده‌بندی، موضوعات اصلی این فصل‌ها را تشکیل می‌دهد. طبیعی است روش‌های ارائه شده در این بخش، برای کار با متن، تطبیق داده شده‌اند. همان‌طور که قبل از این نیز به آن اشاره شد، تحلیل متن ارتباط نزدیکی با حوزه بازیابی اطلاعات دارد. در بخش دوم که تنها شامل فصل نهم کتاب است، به مروری اِجمالی از روش‌های بازیابی اطلاعات از دیدگاه متن‌کاوی پرداخته شده است. فصل‌های دهم تا چهاردهم کتاب، بخش سوم کتاب را تشکیل می‌دهند. در این بخش موضوعات پیشرفته‌ای مانند یادگیری عمیق، استخراج اطلاعات، خلاصه‌سازی، نظرکاوی، تقطیع متن و تشخیص رویداد بررسی می‌شوند.

تمام تلاش خود را انجام داده‌ایم تا ترجمه کتاب به گونه‌ای انجام شود که خوانندگان محترم آن بتوانند مفاهیم آن را به راحتی درک کنند. بدون شک ممکن است برای برخی از واژه‌های انگلیسی بتوان معادل‌های بهتری یافت. آنچه مسلم است این است که شاید گاهی تبدیل واژه‌ها آنچنان که باید و شاید انجام نشده است؛ اما در ادای جملات و بیان موضوع تلاش فراوان شده است تا خوانندگان گرامی با متنی مبهم و گیج کننده روبه‌رو نشوند.

در اینجا لازم می‌دانم از همه اساتید و دانشجویان به خاطر راهنمایی‌های ارزشمندشان در حین آماده-سازی این کتاب سپاسگزاری کنم. همچنین از مدیریت محترم انتشارات آتی‌نگر و دوست عزیزم جناب رامین مولاناپور نیز به خاطر آماده‌سازی، چاپ و پخش این کتاب تشکر می‌کنم. رهین محبت بی‌دریغ خانواده‌ام هستم که با فراهم‌سازی محیطی مناسب مرا یاری نمودند. اما با وجود همه سعی و تلاشی که در تمام مراحل آماده‌سازی این کتاب انجام گرفته است، یقین دارم که عاری از اشتباه نیست، چرا که تنها مکتوب بی‌نقص همان معجزه جاوید قرآن کریم است. در آخر ضمن سپاسگزاری از همه کسانی که مرا یاری داده‌اند و با پذیرش مسئولیت هرگونه کاستی احتمالی، امیدوارم که این اندک مفید افتد.

مهدی اسماعیلی
مردادماه ۹۸