

فهرست مطالب

۱۱	پیشگفتار
۱۵	مقدمه
۱۹	قدردانی‌ها
۲۱	فصل اول: مقدمه
۲۲	۱-۱ هوش مصنوعی، یادگیری ماشین و علوم داده
۲۵	۱-۲ علوم داده چیست؟
۲۵	۱-۲-۱ استخراج الگوهای معنادار
۲۵	۱-۲-۲ ایجاد مدل‌های معرف
۲۶	۱-۲-۳ ترکیب آمار، یادگیری ماشین و رایانش
۲۷	۱-۲-۴ الگوریتم‌های یادگیری
۲۷	۱-۲-۵ زمینه‌های مرتبط
۲۹	۱-۳ دلایل مستدل در رابطه علوم داده
۲۹	۱-۳-۱ حجم
۲۹	۱-۳-۲ ابعاد
۳۰	۱-۳-۳ سؤالات پیچیده
۳۱	۱-۴ طبقه‌بندی علوم داده
۳۳	۱-۵ الگوریتم‌های علوم داده
۳۵	۱-۶ نقشه راه برای این کتاب
۳۵	۱-۶-۱ شروع کار با علوم داده
۳۶	۱-۶-۲ تمرین با استفاده از RapidMiner
۳۶	۱-۶-۳ الگوریتم‌های اصلی
۴۱	منابع

فصل دوم: فرایند علوم داده

۴۳	۲-۱ دانش پیشین
۴۶	۲-۱-۱ هدف
۴۷	۲-۱-۲ حوزه موضوعی
۴۷	۲-۱-۳ داده‌ها
۴۹	۲-۱-۴ علیت در برابر همبستگی

۴۹	۲-۲ آماده‌سازی داده‌ها
۵۰	۲-۲-۱ کاوش داده‌ها
۵۰	۲-۲-۲ کیفیت داده‌ها
۵۱	۲-۲-۳ مقادیر ناموجود
۵۲	۲-۲-۴ انواع داده‌ها و تبدیل
۵۲	۲-۲-۵ تبدیل
۵۲	۲-۲-۶ نقاط پرت
۵۳	۲-۲-۷ انتخاب ویژگی
۵۳	۲-۲-۸ نمونه‌برداری داده‌ها
۵۴	۲-۳ مدل‌سازی
۵۴	۲-۳-۱ مجموعه داده‌های آموزشی و آزمایشی
۵۶	۲-۳-۲ الگوریتم‌های یادگیری
۵۷	۲-۳-۳ ارزیابی مدل
۵۹	۲-۳-۴ مدل‌سازی جمعی
۵۹	۲-۴ کاربرد
۵۹	۲-۴-۱ آمادگی تولید
۶۰	۲-۴-۲ یکپارچه‌سازی فنی
۶۰	۲-۴-۳ زمان پاسخگویی
۶۱	۲-۴-۴ نوسازی مدل
۶۱	۲-۴-۵ همگون‌سازی
۶۱	۲-۵ دانش
۶۲	منابع

فصل سوم: کاوش داده‌ها

۶۵	
۶۶	۳-۱ اهداف کاوش داده‌ها
۶۶	۳-۲ مجموعه داده‌ها
۶۸	۳-۲-۱ انواع داده‌ها
۶۹	۳-۳ آمار توصیفی
۷۰	۳-۳-۱ کاوش تک متغیره
۷۳	۳-۳-۲ کاوش چندمتغیره
۷۵	۳-۴ مصورسازی داده‌ها
۷۶	۳-۴-۱ مصورسازی تک متغیره
۸۱	۳-۴-۲ مصورسازی چندمتغیره
۸۴	۳-۴-۳ مصورسازی داده‌های با ابعاد بالا

۹۰ ۳-۵ نقشه راه برای کاوش داده‌ها
۹۱ منابع

۹۳

فصل چهارم: طبقه‌بندی

۹۴ ۴-۱ درختان تصمیم
۹۴ ۴-۱-۱ نحوه کار آن
۱۰۳ ۴-۱-۲ نحوه پیاده‌سازی
۱۱۶ ۴-۱-۳ نتیجه‌گیری
۱۱۷ ۴-۲ استنتاج قوانین
۱۲۱ ۴-۲-۱ نحوه کار آن
۱۲۴ ۴-۲-۲ نحوه پیاده‌سازی
۱۲۸ ۴-۲-۳ نتیجه‌گیری
۱۲۸ ۴-۳ الگوریتم k- نزدیک‌ترین همسایه
۱۳۰ ۴-۳-۱ نحوه کار آن
۱۳۸ ۴-۳-۲ نحوه پیاده‌سازی
۱۴۱ ۴-۳-۳ نتیجه‌گیری
۱۴۲ ۴-۴ بیز ساده
۱۴۴ ۴-۴-۱ نحوه کار آن
۱۵۳ ۴-۴-۲ نحوه پیاده‌سازی
۱۵۵ ۴-۴-۳ نتیجه‌گیری
۱۵۶ ۴-۵ شبکه‌های عصبی مصنوعی
۱۶۰ ۴-۵-۱ نحوه کار آن
۱۶۳ ۴-۵-۲ نحوه پیاده‌سازی
۱۶۶ ۴-۵-۳ نتیجه‌گیری
۱۶۷ ۴-۶ ماشین‌های بردار پشتیبان
۱۷۱ ۴-۶-۱ نحوه کار آن
۱۷۳ ۴-۶-۲ نحوه پیاده‌سازی
۱۸۰ ۴-۶-۳ نتیجه‌گیری
۱۸۱ ۴-۷ یادگیرنده‌های جمعی
۱۸۳ ۴-۷-۱ نحوه کار آن
۱۸۵ ۴-۷-۲ نحوه پیاده‌سازی
۱۹۴ ۴-۷-۳ نتیجه‌گیری
۱۹۵ منابع

فصل پنجم: روش های رگرسیون

۱۹۹	
۲۰۰	۵-۱ رگرسیون خطی
۲۰۱	۵-۱-۱ نحوه کار آن
۲۰۷	۵-۱-۲ نحوه پیاده سازی
۲۱۴	۵-۱-۳ نقاط واریسی
۲۲۰	۵-۲ رگرسیون لجستیک
۲۲۲	۵-۲-۱ نحوه کار آن
۲۲۸	۵-۲-۲ نحوه پیاده سازی
۲۳۱	۵-۲-۳ خلاصه نکات
۲۳۱	۵-۳ نتیجه گیری
۲۳۳	منابع

فصل ششم: تحلیل وابستگی

۲۳۳	
۲۳۵	۶-۱ قوانین کاوش وابستگی
۲۳۷	۶-۱-۱ مجموعه اقلام
۲۴۰	۶-۱-۲ تولید قانون
۲۴۱	۶-۲ الگوریتم Apriori
۲۴۲	۶-۲-۱ نحوه کار آن
۲۴۶	۶-۳ الگوریتم رشد الگوهای پرتکرار
۲۴۶	۶-۳-۱ نحوه کار آن
۲۵۰	۶-۳-۲ نحوه پیاده سازی
۲۵۵	۶-۴ نتیجه گیری
۲۵۵	منابع

فصل هفتم: خوشه بندی

۲۵۷	
۲۶۳	۷-۱ خوشه بندی K-MEANS
۲۶۵	۷-۱-۱ نحوه کار آن
۲۷۱	۷-۱-۲ نحوه پیاده سازی
۲۷۶	۷-۲ خوشه بندی DBSCAN
۲۷۷	۷-۲-۱ نحوه کار آن
۲۸۱	۷-۲-۲ نحوه پیاده سازی
۲۸۴	۷-۳ نقشه های خودسازمان دهنده
۲۸۶	۷-۳-۱ نحوه کار آن
۲۸۹	۷-۳-۲ نحوه پیاده سازی

منابع ۲۹۷

۲۹۹

فصل هشتم: ارزیابی مدل

۸-۱ ماتریس درهم‌ریختگی ۳۰۱

۸-۲ ROC و AUC ۳۰۳

۸-۳ منحنی‌های ترا ۳۰۶

۸-۴ نحوه پیاده‌سازی ۳۱۰

۸-۵ نتیجه‌گیری ۳۱۵

منابع ۳۱۵

۳۱۷

فصل نهم: متن‌کاوی

۹-۱ نحوه کار آن ۳۲۰

۹-۱-۱ تکرار عبارت - معکوس تکرار در سند ۳۲۰

۹-۱-۲ مجموعه اصطلاحات ۳۲۲

۹-۲ نحوه پیاده‌سازی ۳۲۷

۹-۲-۱ پیاده‌سازی ۱: خوشه‌بندی کلمات کلیدی ۳۲۷

۹-۲-۲ پیاده‌سازی ۲: پیش‌بینی جنسیت نویسندگان وبلاگ ۳۳۲

۹-۳ نتیجه‌گیری ۳۴۱

منابع ۳۴۲

۳۴۳

فصل دهم: یادگیری ژرف

۱۰-۱ زمستان هوش مصنوعی ۳۴۷

۱۰-۲ نحوه کار آن ۳۵۳

۱۰-۲-۱ مدل‌های رگرسیون به‌عنوان شبکه‌های عصبی ۳۵۴

۱۰-۲-۲ گرادیان نزولی ۳۵۶

۱۰-۲-۳ نیاز به پس‌انتشار ۳۶۰

۱۰-۲-۴ طبقه‌بندی بیش از ۲ دسته: تابع بیشینه هموار (Softmax) ۳۶۱

۱۰-۲-۵ شبکه‌های عصبی کانولوشن ۳۶۳

۱۰-۲-۶ لایه متراکم ۳۷۰

۱۰-۲-۷ لایه حذف تصادفی ۳۷۱

۱۰-۲-۸ شبکه‌های عصبی تراجعی ۳۷۱

۱۰-۲-۹ خودرمزگذارها ۳۷۳

۱۰-۲-۱۰ مدل‌های هوش مصنوعی مرتبط ۳۷۴

۱۰-۳ نحوه پیاده‌سازی ۳۷۵

۳۸۱	۱۰-۴ نتیجه گیری
۳۸۱	منابع

۳۸۳

فصل یازدهم: موتورهای پیشنهاددهنده

۳۸۷	۱۱-۱ مفاهیم موتور پیشنهاددهنده
۳۹۲	۱۱-۱-۱ انواع موتورهای پیشنهاددهنده
۳۹۵	۱۱-۲ پالایش مشارکتی
۳۹۶	۱۱-۲-۱ روش های مبتنی بر همسایگی
۴۰۹	۱۱-۲-۲ تجزیه عاملی ماتریس
۴۱۶	۱۱-۳ پالایش مبتنی بر محتوا
۴۱۹	۱۱-۳-۱ محاسبه نمایه کاربر
۴۲۷	۱۱-۳-۲ مدل های یادگیری با ناظر
۴۳۳	۱۱-۴ پیشنهاددهنده های ترکیبی
۴۳۵	۱۱-۵ نتیجه گیری
۴۳۸	منابع

۴۴۱

فصل دوازدهم: پیش بینی سری های زمانی

۴۴۶	۱۲-۱ تجزیه سری های زمانی
۴۴۹	۱۲-۱-۱ تجزیه کلاسیک
۴۵۰	۱۲-۱-۲ نحوه پیاده سازی
۴۵۳	۱۲-۲ روش مبتنی بر هموارسازی
۴۵۳	۱۲-۲-۱ روش های پیش بینی ساده
۴۵۵	۱۲-۲-۲ هموارسازی نمایی
۴۵۸	۱۲-۲-۳ نحوه پیاده سازی
۴۶۰	۱۲-۳ روش های مبتنی بر رگرسیون
۴۶۱	۱۲-۳-۱ رگرسیون
۴۶۲	۱۲-۳-۲ رگرسیون با تغییرات فصلی
۴۶۵	۱۲-۳-۳ میانگین متحرک یکپارچه خودهمبسته
۴۷۲	۱۲-۳-۴ ARIMA فصلی
۴۷۵	۱۲-۴ روش های یادگیری ماشین
۴۷۶	۱۲-۴-۱ پنجره بندی
۴۸۲	۱۲-۴-۲ شبکه عصبی خودهمبسته
۴۸۵	۱۲-۵ ارزیابی عملکرد
۴۸۵	۱۲-۵-۱ مجموعه داده اعتبارسنجی

۴۸۸	۱۳-۵-۲ اغراض سطحی پنجره کشویی
۴۸۸	۱۳-۶ نتیجه گیری
۴۸۹	۱۳-۶-۱ پیش بینی به روش ها
۴۹۰	منابع

۴۹۱ فصل سیزدهم: تشخیص ناهنجاری

۴۹۲	۱۳-۱ مفاهیم
۴۹۲	۱۳-۱-۱ علل نقاط برت
۴۹۵	۱۳-۱-۲ تکبیک های تشخیص ناهنجاری
۴۹۸	۱۳-۲ تشخیص نقطه برت بر اساس فاصله
۴۹۹	۱۳-۲-۱ نحوه کار آن
۴۹۹	۱۳-۲-۲ نحوه پیاده سازی
۵۰۳	۱۳-۳ تشخیص نقطه برت بر اساس جدالی
۵۰۳	۱۳-۳-۱ نحوه کار آن
۵۰۳	۱۳-۳-۲ نحوه پیاده سازی
۵۰۶	۱۳-۴ عامل برت محلی
۵۰۶	۱۳-۴-۱ نحوه کار آن
۵۰۷	۱۳-۴-۲ نحوه پیاده سازی
۵۰۹	۱۳-۵ نتیجه گیری
۵۱۰	منابع

۵۱۱ فصل چهاردهم: انتخاب ویژگی

۵۱۳	۱۴-۱ طبقه بندی روش های انتخاب ویژگی
۵۱۴	۱۴-۲ تحلیل مؤلفه اصلی
۵۱۵	۱۴-۲-۱ نحوه کار آن
۵۱۷	۱۴-۲-۲ نحوه پیاده سازی
۵۲۲	۱۴-۳ پالایش مبتنی بر نظریه اطلاعات
۵۲۴	۱۴-۴ پالایش مبتنی بر خی دو
۵۲۷	۱۴-۵ انتخاب ویژگی نوع روکش
۵۲۹	۱۴-۵-۱ حذف پسرو
۵۳۲	۱۴-۶ نتیجه گیری
۵۳۳	منابع

فصل پانزدهم: شروع کار با RapidMiner

۵۳۵

- ۱-۱۵ رابط کاربر و مجموعه اصطلاحات..... ۵۳۶
- ۲-۱۵ ابزارهای وارد کردن و خروجی گرفتن از داده‌ها..... ۵۴۲
- ۳-۱۵ ابزارهای مصورسازی داده‌ها..... ۵۴۵
- ۴-۱۵ ابزارهای تبدیل داده..... ۵۴۸
- ۵-۱۵ ابزارهای نمونه‌برداری و مقدار ناموجود..... ۵۵۳
- ۶-۱۵ ابزارهای بهینه‌سازی..... ۵۵۷
- ۷-۱۵ یکپارچگی با R..... ۵۶۴
- ۸-۱۵ نتیجه‌گیری..... ۵۶۵
- منابع..... ۵۶۵

مقایسه الگوریتم‌های علوم داده‌ها

۵۶۷

- رگرسیون: پیش‌بینی یک متغیر هدف کمی رسته‌ای..... ۵۶۷
- رگرسیون: پیش‌بینی یک متغیر هدف کمی عددی..... ۵۶۹
- تحلیل وابستگی: فرایند بدون ناظر برای یافتن روابط بین ارقام..... ۵۶۹
- خوشه‌بندی: فرایندی بدون ناظر برای یافتن گروه‌های با معنی در داده‌ها..... ۵۷۰
- تشخیص ناهنجاری: تکنیک‌های با ناظر و بدون ناظر برای یافتن نقاط پرت در داده‌ها..... ۵۷۱
- یادگیری ژرف: آموزش با استفاده از لایه‌های متعدد بازنمایی داده‌ها..... ۵۷۲
- پیشنهاددهنده‌ها: یافتن ارجحیت کاربر برای یک قلم..... ۵۷۳
- پیش‌بینی سری‌های زمانی: پیش‌بینی مقدار آتی یک متغیر..... ۵۷۴
- انتخاب ویژگی: انتخاب مهم‌ترین صفات خاصه..... ۵۷۵