

فهرست

- پیشگفتار..... ۵
- چرا این کتاب را می‌نویسم؟..... ۵
- چه انگیزه‌ای مرا برای نوشتن این کتاب ترغیب کرد؟..... ۵
- خوانندگان کتاب چه کسانی هستند؟..... ۶
- چه چیزی در کتاب خواهید یافت؟..... ۶
- بعد از خواندن کتاب چه تأثیری خواهد داشت؟..... ۷

فصل ۱: جنایت بوستون..... ۱۷

- مقدمه..... ۱۷
- ساختار..... ۱۸
- هدف..... ۱۸
- داده چیست؟..... ۱۸
- انواع داده‌ها..... ۱۸
- بیاید در مورد مجموعه داده بوستون صحبت کنیم..... ۱۹
- فرهنگ لغت داده..... ۱۹
- چارچوب O.S.E.M.N..... ۲۱
- به دست آوردن داده چیست؟..... ۲۱
- پاکسازی داده‌ها چیست؟..... ۲۲
- یافتن انواع داده‌ها..... ۲۲
- چگونه داده‌های از دست رفته را مدیریت کنیم؟..... ۲۴
- با مقادیر تکراری چگونه رفتار کنیم؟..... ۲۷

۳۰.....	کاوش داده چیست؟
۳۲.....	تجزیه و تحلیل ویژگی تک متغیره.....
۳۲.....	توزیع آماری چیست؟
۳۵.....	تجزیه و تحلیل ویژگی های چند متغیره.....
۴۷.....	فصل ۲: گزارش شادی جهانی
۴۷.....	مقدمه
۴۷.....	ساختار
۴۸.....	هدف
۴۸.....	بیاید در مورد مجموعه داده صحبت کنیم
۴۸.....	پیش نیازها
۴۸.....	یک جمع بندی سریع
۵۱.....	کاوش داده ها
۵۱.....	انواع مختلف داده
۵۱.....	داده های بازنمایی شده در مقیاس اسمی (دسته بندی شده)
۵۲.....	داده های بازنمایی شده در مقیاس ترتیبی
۵۲.....	اندازه گیری داده ها
۵۲.....	متغیرهای گسسته
۵۲.....	متغیرهای پیوسته
۵۳.....	مقیاس فاصله
۵۳.....	مقیاس دایره ای
۵۳.....	مقیاس نسبت
۵۳.....	توزیع ها
۵۶.....	توزیع نرمال
۵۸.....	«قاعده تجربی» در آمار چیست؟
۵۹.....	نمودار توزیع
۶۱.....	نخمن چگالی کرنل
۶۲.....	نمودار جمعی ای
۶۶.....	توزیع چوله
۶۷.....	چوله به چپ (انحراف منفی)

۶۸ "سلامتی (امید به زندگی)" چیست؟
۶۸ نمودار توزیع
۷۰ برآورد چگالی کرنل
۷۲ نمودار جعبه
۷۳ چوله به راست (انحراف مثبت)
۷۳ نمودار توزیع
۷۴ برآورد چگالی کرنل
۷۹ نمودار جعبه
۸۰ سایر توزیع‌ها
۸۳ فصل ۳: گونه‌های زنبق
۸۳ مقدمه
۸۴ ساختار
۸۴ هدف
۸۴ مقدمه‌ای بر یادگیری ماشین
۸۵ یادگیری ماشین چیست ^۱ ؟
۸۶ انواع یادگیری ماشینی
۸۶ یادگیری نظارت شده
۸۶ یادگیری بدون نظارت
۸۶ یادگیری تقویتی
۸۷ خط لوله یادگیری ماشین
۸۸ درک مجموعه داده
۸۹ تجزیه و تحلیل داده‌های اکتشافی
۸۹ بارگذاری مجموعه داده
۹۰ Scikit-Learn
۹۲ Seaborn
۹۳ تحلیل داده‌ها
۹۳ نمودار جفتی
۹۴ مشاهدات
۹۵ همبستگی

۹۹.....	چگونه الگوریتم‌های ML را انتخاب کنیم؟
۹۹.....	طبقه‌بندی.....
۱۰۰.....	K - نزدیکترین همسایگان (K-nn).....
۱۰۱.....	چگونه کار می‌کند؟
۱۰۲.....	آموزش مدل.....
۱۰۳.....	ارزیابی.....
۱۰۳.....	ماتریس سردرگمی چیست؟
۱۰۵.....	تنظیم مدل.....
۱۰۷.....	خوشه‌بندی.....
۱۰۸.....	K-means.....
۱۰۸.....	چگونه کار می‌کند؟
۱۱۰.....	آموزش مدل.....
۱۱۵.....	ارزیابی و تنظیم مدل.....
۱۱۵.....	اینرسی یا معیار مجموع مربعات درون خوشه ای.....
۱۱۵.....	روش آرنج با استفاده از اینرسی.....
۱۱۸.....	امتیاز سیلوئت ^{۳۰}
۱۲۱.....	رگرسیون.....
۱۲۴.....	رگرسیون خطی.....
۱۲۴.....	چگونه کار می‌کند؟
۱۲۵.....	آموزش مدل.....
۱۳۰.....	ارزیابی.....
۱۳۱.....	میانگین خطای مطلق.....
۱۳۱.....	خطای میانگین مربعات.....
۱۳۱.....	ریشه میانگین مربعات خطا ^{۴۲}
۱۳۲.....	امتیاز R2 یا ضریب تعیین.....
۱۳۹.....	فصل ۴: تشخیص تقلب در کارت اعتباری.....
۱۳۹.....	مقدمه.....
۱۴۰.....	ساختار.....
۱۴۰.....	هدف.....

۱۴۰	بیاید داده‌ها را درک کنیم
۱۴۱	منبع مجموعه داده:
۱۴۱	توضیحات ستون
۱۴۲	مجموعه داده نامتعادل چیست؟
۱۴۴	شناخت ویژگی‌ها
۱۴۸	پیش نیازها
۱۴۸	نرمال‌سازی مقیاس‌بندی ویژگی
۱۵۰	مقیاس‌بندی حداقل/حداکثر ویژگی (مقیاس مجدد)
۱۵۱	نرمال‌سازی میانگین
۱۵۱	استانداردسازی (نرمال‌سازی امتیاز Z-score)
۱۵۱	تجزیه و تحلیل مؤلفه اصلی
۱۵۲	اعتبار سنجی متقابل
۱۵۴	تحلیل داده‌ها
۱۶۶	مقیاس بندی
۱۶۶	مقیاس کننده استاندارد
۱۶۹	مقیاس‌بندی مقاوم
۱۷۰	ترانسفورماتور قدرت
۱۷۱	ترانسفورماتور کوانتیل
۱۷۶	تقسیم مجموعه داده
۱۷۹	مدیریت عدم تعادل
۱۸۰	نمونه فرعی چیست؟
۱۸۰	نمونه‌گیری مجدد چیست؟
۱۸۳	نمونه‌گیری کم تصادفی
۱۸۷	نمونه‌برداری تصادفی بیش از حد
۱۸۹	پیوندهای توکم برای نمونه‌برداری کم
۱۹۳	تکنیک نمونه‌برداری بیش از حد از اقلیت مصنوعی
۱۹۶	تجزیه و تحلیل مجدد داده‌ها
۱۹۶	ویژگی مثبت (V1..., V28)
۱۹۷	داده‌های آموزشی اصلی

۱۹۷ نمونه‌گیری بیش از حد تصادفی
۱۹۸ نمونه‌گیری بیش از حد تصادفی
۱۹۹ نمونه‌برداری کم پیوند تو مک
۱۹۹ SMOTE
۲۰۰ ویژگی منفی (V1.. , V28)
۲۰۱ ویژگی‌های مقیاس شده
۲۰۲ ویژگی Time
۲۰۳ ویژگی Amount
۲۰۵ مدل‌سازی
۲۰۵ الگوریتم‌های یادگیری ماشین
۲۰۶ رگرسیون لجستیک
۲۰۶ چرا از رگرسیون لجستیک (مقایسه با رگرسیون خطی) استفاده کنیم؟
۲۰۶ تابع لاجیت چیست؟
۲۰۷ مشتق:
۲۰۹ ساختار درخت تصمیم
۲۰۹ الگوریتم
۲۱۰ ریاضیات پشت ASM
۲۱۱ ماشین بردار پشتیبان (SVM)
۲۱۱ چگونه کار می‌کند؟
 ریاضیات پایه در پشت طبقه‌بندی‌کننده حاشیه سخت خطی (SVM خطی) برای موارد
۲۱۴ کاملاً قابل تفکیک:
۲۱۵ آموزش مدل
۲۱۵ آماده‌سازی داده‌ها
۲۱۶ تله متریک
۲۱۷ آموزش و ارزشیابی
۲۳۳ فصل ۵: بیماری قلبی UCI
۲۳۳ مقدمه
۲۳۴ ساختار
۲۳۴ هدف

۲۳۴	پیش نیازها
۲۳۴	چرا یادگیری ماشین در پزشکی بسیار مهم است؟
۲۳۶	هوش مصنوعی قابل توضیح چیست؟
۲۳۷	منظم سازی
۲۳۷	LASSO (حداقل انقباض مطلق و اپراتور انتخاب) ^۲
۲۳۸	رگرسیون ستیغی ^۷
۲۳۹	آزمایش فرضیه
۲۴۱	روش گروهی
۲۴۲	کیسه‌ای
۲۴۲	تقویت
۲۴۶	تحلیل داده‌ها
۲۴۷	همبستگی
۲۴۹	ویژگی «ca»
۲۵۲	ویژگی «oldpeak»
۲۵۴	ویژگی «thalach»
۲۵۶	ویژگی «cp»
۲۵۸	تقسیم مجموعه داده
۲۵۹	مدلسازی یادگیری ماشینی
۲۵۹	الگوریتم‌های یادگیری ماشین
۲۵۹	طبقه‌بندی کننده نایو بیز ^{۳۱} و ^{۳۲}
۲۶۱	جنگل تصادفی
۲۶۳	آموزش مدل
۲۶۴	آموزش و ارزشیابی پایه
۲۶۴	نایو بیز
۲۶۶	روش مجموعه (دسته بندی)
۲۶۹	روش گروهی (تقویت کننده)
۲۷۱	جنگل تصادفی
۲۷۴	انتخاب ویژگی و جستجوی شبکه
۲۷۴	انتخاب ویژگی

۲۷۵	خط لوله
۲۷۸	جستجوی شبکه
۲۸۳	هوش مصنوعی قابل توضیح
۲۸۳	اهمیت ویژگی
۲۸۵	تجسم مدل
۲۸۹	طرح وابستگی جزئی
۲۹۱	SHAP (توضیحات افزودنی شپ)
۲۹۷	منابع