

۱۵۱۷۶۷۴

داده کاوی در شبکه‌های اجتماعی آنلاین

نرگس زاوندیان



تهران - ۱۳۹۶

377 7761

کتابخانه

کتابخانه مرکزی و اسناد و کتابخانه ملی

www.ketab.ir

شابک	978-600-5770-44-5 :
شماره کتابشناسی ملی	۴۷۶۴۰۳۰ :
عنوان و نام پدیدآور	داده کاوی در شبکه های اجتماعی، ملین آراس ندای آریان :
مشخصات نشر	تهران : ندای آریانا، ۱۳۹۶.
مشخصات ظاهری	۱۰۰ ص.
موضوع	داده کاوی :
موضوع	Data mining :
موضوع	شبکه های اجتماعی پیوسته :
موضوع	Online social networks :
رده بندی دیویی	۳۱۲/۰۰۶ :
رده بندی کنگره	۹۷۶QA ۹۱۳۹۶ ن ۵۲/د :

بسم الله الرحمن الرحيم



داده کاوی در شبکه‌های اجتماعی آنلاین

تالیف: نرگس نهاوندیان

ناشر: ندای آریانا

چاپ: اول ۱۳۹۶

شمارگان: ۵۰۰ نسخه

شابک: ۹۷۸-۶۰۰-۵۷۷۰-۴۴-۵

قیمت ۱۰۰۰۰ تومان

کلیه حقوق برای ناشر محفوظ است

پیشگفتار

شبکه‌های اجتماعی، بویژه شبکه‌های اجتماعی آنلاین (OSNs) محبوبیت بزرگی در سال‌های اخیر کسب کرده‌اند و در طی این سالها یکی از مهم‌ترین پدیده‌های علوم اجتماعی و کامپیوتر به شمار می‌روند. در این پایان‌نامه به منظور بررسی عوامل مؤثر داده کاوی در شبکه‌های اجتماعی از چهار مدل داده کاوی (شامل رگرسیون لجستیک، درخت تصمیم، شبکه‌ی عصبی و مدل‌های - لوجستیک تصادفی) استفاده شده است. سپس با بررسی کارکرد مدلها، در هر دو مدل داده‌های مرتبط با داده کاوی، به اندازه‌گیری خطاهای پیش بینی این چهار روش پرداخته شده است. در حالت کلی مقایسه‌ها در دو حالت انجام شده است. در حالت اول هر کدام از روش‌ها بدون استفاده از نمونه‌برداری پیاده‌سازی شده و در حالت دوم با استفاده از نمونه‌برداری عمل شده است. با توجه به نتایج به دست آمده و علم به اینکه پارامتر دقت کلی مدل، مهمترین پارامتر برای ارزیابی این روش‌ها است، روش رگرسیون لجستیک در هر دو حالت دارای بالاترین دقت بود و گزینه مناسب برای پیش‌بینی داده کاوی در شبکه‌های آنلاین می‌باشد.

فهرست مطالب

عنوان	شماره صفحه
فصل اول	۱۵
۱-۱ انگیزه	۱۹
۱-۲ بیان مسئله	۲۱
۱-۳ هدف	۲۲
۱-۴ تعریف اصطلاحات	۲۳
۱-۵ چهار چوب پایان نامه	۲۳
فصل دوم	۲۵
۱-۶ داده کاوی	۲۷
۱-۶-۱ مراحل داده کاوی	۲۸
۱-۷ یادگیری ماشین	۳۱
۱-۸ الگوریتمهای طبقه بندی	۳۲
۱-۸-۱ درخت تصمیم	۳۳
۱-۸-۲ جنگلهای تصادفی	۳۴
۱-۸-۳ رگرسیون منطقی	۳۷
۱-۸-۴ شبکههای عصبی	۳۹
۱-۹ تکنیکهای ارزیابی مدل	۴۲
۱-۹-۱ ماتریس خطا برای طبقه بندی دو کلاسه	۴۳
۱-۹-۲ ROC	۴۴
۱-۹-۳ منحنی ROC	۴۷
۱-۱۰ داده کاوی آموزشی	۴۹

فصل سوم ۵۱

فصل چهارم ۵۷

۱-۱۱ داده‌های ورودی ۶۰

۱-۱۲ مدل‌سازی پیش‌گویانه ۶۳

۱-۱۲-۱ پیش‌بینی با استفاده از درخت تصمیم بدون استفاده از نمونه‌برداری ۶۳

۱-۱۲-۲ پیش‌بینی با استفاده از درخت تصمیم با استفاده از نمونه‌برداری ۶۶

۱-۱۲-۳ پیش‌بینی با استفاده از رگرسیون لجستیک بدون استفاده از نمونه‌برداری ۶۸

۱-۱۲-۴ پیش‌بینی با استفاده از رگرسیون لجستیک با استفاده از نمونه‌برداری ۷۰

۱-۱۲-۵ پیش‌بینی با استفاده از شبکه عصبی بدون استفاده از نمونه‌برداری ۷۲

۱-۱۲-۶ پیش‌بینی با استفاده از شبکه عصبی با استفاده از نمونه‌برداری ۷۵

۱-۱۲-۷ پیش‌بینی با استفاده از جنگل‌های تصادفی بدون استفاده از نمونه‌برداری ۷۸

۱-۱۲-۸ پیش‌بینی با استفاده از جنگل‌های تصادفی با استفاده از نمونه‌برداری ۸۱

۱-۱۳ نتیجه‌گیری ۸۳

فصل پنجم ۸۵

۱-۱۴ مقایسه روش درخت تصمیم ۸۷

۱-۱۵ مقایسه روش رگرسیون لجستیک ۸۸

۱-۱۶ مقایسه روش شبکه عصبی ۸۹

۱-۱۷ مقایسه روش جنگل‌های تصادفی ۹۰

۱-۱۸ مقایسه کلی روش‌ها بدون استفاده از نمونه‌برداری ۹۱

۱-۱۹ مقایسه کلی روش‌ها با استفاده از نمونه‌برداری ۹۳

۱-۲۰ نتیجه‌گیری ۹۵

مراجع mark not defined