

۲۱۲۰۲۵۱

نگینک های داده کاوی در

تشخیص ایمییل های اسپم

مؤلفین: مهندس سکینه شهریاری
مهندس سید مصطفی امامی



انتشارات کتاب دانش

۹۹ هزار

www.ketab.ir

سوشناسه	شهریاری، سکینه، ۱۳۶۳-
عنوان و نام پدیدآور	تکنیک‌های داده‌کاوی در تشخیص ایمیل‌های اسپم/مؤلفین سکینه شهریاری، سیدمصطفی امامی.
مشخصات نشر	تهران: آریا دانش، ۱۳۹۹.
مشخصات ظاهری	۱۲۳ ص. مصور، جدول، نمودار.
شابک	۲۲۰۰۰ ریال ۶۸۶۶۵۵-۶۲۲-۹۷۸:
وضعیت فهرست نویسی	فیا
یادداشت	واژه‌نامه.
یادداشت	کتابنامه: ص. ۱۱۰ - ۱۱۷.
موضوع	هرنامه (پست الکترونیکی)
موضوع	Spam (Electronic mail)
موضوع	هرنامه (پست الکترونیکی) - پالایش
موضوع	Spam filtering (Electronic mail)
موضوع	هرنامه (پست الکترونیکی) - پیشگیری
موضوع	Prevention — (Electronic mail) Spam
موضوع	داوی
موضوع	Data mining
شناسه افزوده	می، لمد، می، ۱۳۶۷-
رده بندی کنگره	۷۳۳، ۴۵۱ د
رده بندی دیویی	۹۲/۰۰۴
شماره کتبی‌شناسی ملی	۶۱۹۵۲۴۵

هرگونه تکثیر کامل یا قسمتی از کتاب بدون اجازت پدیدآورنده یا ناشر خلاف قانون، شرع و اخلاق است. موارد تخلف را به دفتر مرکزی "انف" گزارش فرمائید.

عنوان: تکنیک‌های داده‌کاوی در تشخیص ایمیل‌های اسپم
مؤلفین: مهندس سکینه شهریاری، مهندس سید مصطفی امامی



۱. (مدرس مرکز علمی کاربردی شهرداری شیانکاره) sh_hrv_193us@gmail.com

ناشر: آریا دانش

صفحه آرا: سکینه شهریاری

نوبت: اول خرداد ۱۳۹۹

شمارگان: ۱۰ نسخه شابک: ۶۸۶۶-۶۲۲-۹۷۸

قیمت: ۲۲۰۰۰ تومان

مرکز پخش: تهران، میدان انقلاب، خیابان آزادی، خیابان بهزاد، پلاک ۳۲ واحد ۱۴ تلفن ۸۸۵۶۲۲۲۳

حق چاپ برای مؤلف و ناشر محفوظ است

فصل اول - مقدمه ۱۲

۱-۱ مقدمه ۱۲

۱-۱ بیان مسئله ۱۴

۱-۱ اهداف ۱۶

فصل دوم - مفاهیم اولیه روش‌های رایج ۲۰

۱-۲ الگوریتم‌های یادگیری مبتنی بر مثال ۲۰

۱-۲-۱ الگوریتم رتبه‌بندی مجموعه آموزشی ۲۰

۱-۲-۲ الگوریتم SVM ۲۶

۱-۲-۳ کاهش داده‌های آموزشی بر اساس تشخیص لبه ۲۸

۱-۲-۴ خوشه‌بندی Medoidshift ۳۵

۱-۲-۵ الگوریتم Crisp ۳۹

۱-۲-۶ الگوریتم CBP ۴۱

۲-۲ مروری بر پژوهش‌های انجام‌شده در گذشته ۵۲

۲-۲-۱ طبقه‌بندی کلاس‌های اسپم و غیر اسپم ۵۲

۲-۲-۲ دیگر اهداف پژوهش تجزیه و تحلیل داده‌های ایمیل ۵۴

۲-۲-۳ طبقه‌بندی هستی‌شناسی محتویات ایمیل‌ها ۵۵

۲-۲-۴ روش‌های داده‌کاوی برای شناسایی اسپم ۵۶

۲-۲-۵ معیارهای ارزیابی الگوریتم‌های تشخیص اسپم ۶۶

فصل سوم - سیستم پیشنهادی و معرفی روش‌های بکاررفته ۷۰

۱-۳ الگوریتم‌های بکاررفته در روش پیشنهادی ۷۰

۱-۳-۱ الگوریتم تشخیص لبه ۷۱

۱-۳-۲ الگوریتم k-means ۷۱

۷۲	۳-۱-۳ medoidshift الگوریتم
۷۲	۴-۱-۳ SVM الگوریتم
۷۴	۲-۳ روش پیشنهادی
۷۶	۱-۲-۳ مراحل اجرای روش پیشنهادی
۷۷	۲-۲-۳ پارامتر کنترلی rate
۸۲	فصل چهارم- تحلیل نتایج
۸۳	۱-۱ نتایج مربوط به داده‌های آزمایشی
۸۳	۱-۱-۴ آزمایش مجموعه داده شطرنجی
۸۸	۲-۱-۴ آزمون ایش بر روی مجموعه داده Gaussian random binary data
۹۲	۲-۲-۴ آزمون مرتبط به مجموعه داده‌های واقعی
۹۲	۲-۲-۴ تشخیص ایمیل‌های اسپم با استفاده از روش ترکیبی پیشنهادی
۹۴	۲-۲-۴ آزمون سوره‌های اصلاحی با روش‌های دیگر
۹۵	۱-۲-۲-۴ نرخ درستی الگوریتم‌ها
۹۷	۲-۲-۲-۴ نرخ کاهش الگوریتم‌ها
۹۹	۳-۴ سایر آزمایش‌های صورت گرفته
۹۹	۱-۳-۴ تابع ابتکاری جهت کاهش
۱۰۶	فصل پنجم- نتیجه‌گیری و پیشنهادات آینده
۱۰۶	۱-۵ نتیجه‌گیری
۱۰۸	۲-۵ پیشنهادات و کارهای آینده
۱۱۰	منابع و مؤخذ

فهرست جدول‌ها

صفحه

عنوان

۸۲	فصل چهارم - تحلیل نتایج
۸۷	جدول ۱-۴ نتایج حاصل از اجرای هر کدام از الگوریتم‌ها
۸۸	جدول ۲-۴ تأثیر مقادیر متفاوت rate بر سایر پارامترها
۹۳	جدول ۳-۴ مقایسه نتایج الگوریتم پیشنهادی و سایر الگوریتم‌ها به ازای $rate = 0.003$
۹۴	جدول ۴-۴ مشخصات مجموعه داده‌های استفاده شده در آزمایش‌های
۹۶	جدول ۵-۴ مقایسه نرخ درستی روش پیشنهادی و دیگر روش‌ها
۹۸	جدول ۶-۴ مقایسه نرخ کاهش روش پیشنهادی و دیگر روش‌ها