



پایان‌نامه‌ی کارشناسی ارشد: فاطمه سادات جمالی، ۱۳۹۶

## تشخیص ایمیل‌های اسپم از غیر اسپم با استفاده از داده کاوی

اخیرا با توسعه‌ی فناوری‌های نوین و ابزارهای ارتباطی سریع مانند اینترنت، پست الکترونیک، بعنوان یکی از مهم‌ترین این ابزارها برای ارسال و دریافت اطلاعات، بشمار می‌رود و در این راستا، ارزان بودن این تکنولوژی باعث شده تا بستر مناسبی برای افراد سودجو در زمینه مقاصد مختلفی مانند تبلیغ محصولات، ترویج عقاید، ارسال انواع بدافزارها و ویروس‌ها برای تخریب سیستم کاربران و یا نفوذ به اطلاعات محرمانه آنها و ... را فراهم نماید. همچنین با توجه به رشد سریع شبکه‌های اجتماعی و نامعلوم بودن هویت ارسال کنندگان ایمیل‌ها، کاربران در سطح وسیعی درگیر یکسری ایمیل‌های اسپم یا به اصطلاح هرزنامه می‌شوند که سبب اتلاف پهنای باند شبکه و صرف زمان و هزینه برای کاربران می‌شود، بهمین دلیل، برقراری امنیت کامپیوترها بدلیل حملات انجام شده توسط این هرزنامه‌ها بسیار ضروری می‌باشد. روش سنتی تشخیص هرزنامه بر اساس امضا بوده که با توجه به ترفندهای جدید سویی‌استفاده کنندگان، موثر نیست و امروزه تشخیص هرزنامه و مشکلات دسته‌بندی ایمیل‌ها با استفاده از سیستم یادگیری ماشین و داده‌کاوی حل شده است. در روش ذکر شده، بایستی هرزنامه از غیر هرزنامه را تشخیص داده و سپس دسته بندی ایمیل‌ها را انجام دهیم تا نرخ خطا کاهش یابد و پس از آن، فیلتر کردن هرزنامه‌ها در سطوح مختلف را می‌توان اعمال نمود. دسته بندی بر اساس یک سری ویژگی انجام می‌گیرد که انتخاب ویژگی‌های مناسب، تاثیر زیادی بر روی کارایی دسته بند می‌گذارد. با توجه به اینکه کاهش ویژگی می‌تواند به دو صورت استخراج ویژگی و انتخاب ویژگی، بر اساس معیار خاص باشد، در اینمقاله ضمن آشنایی با روش‌های مختلف تشخیص و جلوگیری از هرزنامه‌ها، به روشی مناسب برای دسته بندی هرزنامه‌ها با استفاده از تکنیک‌های به روز داده‌کاوی می‌پردازیم و همچنین روشی برای استخراج ویژگی‌های مناسب بمنظور کاهش حجم دیتاست‌ها و افزایش سرعت عمل بیشتر برای دسته بندی هرزنامه‌ها، ارائه می‌شود.

**کلیدواژه‌ها:** هرزنامه، دسته بندی، کاهش ویژگی، استخراج ویژگی، داده‌کاوی

شماره‌ی پایان‌نامه: ۱۲۷۴۱۰۰۶۹۵۲۰۰۲

تاریخ دفاع: ۱۳۹۶/۰۴/۲۸

رشته‌ی تحصیلی: مهندسی کامپیوتر- نرم افزار

دانشکده: فنی و مهندسی

استاد راهنما: مهندس یاسر علمی سولا

استاد مشاور: مهندس حسام حسن پور

**M.A. Thesis:**



## Distinguish spam from non-spam emails using data mining

Recently, with the development of new technologies and fast communication tools such as the Internet ,e-mail it is considered as one of the most important tools for sending and receiving information. In this regard, the cheapness of this technology has made the bed It is suitable for people who are profitable for various purposes, such as promoting products, promoting opinions, sending various types of malware and viruses to destroy users' systems or penetrating their confidential information, etc. Also, due to the rapid growth of social networks and the uncertainty of identity of the senders of emails, users are widely involved in a series of spam or spam emails that cause waste of network bandwidth and spend time and money for For users, it's very necessary to keep the computers safe because of attacks by these spam. The traditional signature-based signature-based spam detection method is not effective due to new user tricks. Today, spam detection and mail ordering problems are solved using a machine learning and data mining system. In the method described, we need to detect spam from non-spam and then sort the emails to reduce the error rate, and then filter spam at various levels can be applied. Based on a series of features, selecting the appropriate features will greatly affect the functionality of the category. Considering that the reduction of the attribute can be in two ways: extracting the attribute and selecting the attribute, based on a specific criterion, in this category, along with familiarization with different methods of detection and prevention of spam, in a proper way for categorizing spam Using modern data-driven techniques, we present a way to extract the proper features to reduce the volume of data and increase the speed of the operation of spam categories.