

## عنوان مقاله:

روشی مبتنی بر تصفیه برای انتخاب ویژگی از داده های ابعاد بالا با هدف دسته بندی اسناد

## محل انتشار:

دومین کنفرانس ملی رویکردهای نوین در مهندسی کامپیوتر و برق (سال: 1395)

تعداد صفحات اصل مقاله: 5

## نویسندگان:

مریم کرباسی - کارشناسی ارشد کامپیوتر، دانشگاه علوم و تحقیقات واحد نیشابور

محمدحسین معطر - استادیار گروه کامپیوتر، دانشگاه آزاد اسلامی واحد مشهد

## خلاصه مقاله:

با توجه به اینکه دسته بندی اسناد منجر به بازیابی اطلاعات مناسب می شود ولی ابعاد بالای فضای ویژگی ها یک چالش مهم در کارایی دسته بندی اسناد است. این چالش موجب کاهش صحت و پیچیدگی بالای محاسبات در الگوریتم های یادگیری ماشین جهت دسته بندی اسناد می شود. در این مقاله یک روش دو مرحله ای مبتنی بر تصفیه، جهت کاهش ابعاد با هدف حفظ دقت دسته بندی اسناد ارائه داده ایم. در فاز تصفیه ویژگی ها توسط معیار IG امتیازدهی می شوند و مجموعه بهینه از ویژگی ها را انتخاب کرده ایم. سپس در فاز بعدی، ویژگی های مازاد را از مجموعه بهینه ویژگی ها حذف کرده ایم. در نهایت اسناد به همراه بهترین ویژگی های انتخاب شده، توسط دسته بندی های Naïve Bayes و c4.5 دسته بندی می شوند. با توجه به آزمایشات انجام شده بر روی مجموعه داده روبرتز، میزان صحت روش پیشنهادی با دسته بند c4.5؛ 0.927 است.

## کلمات کلیدی:

دسته بندی اسناد، انتخاب ویژگی، معیار IG، روش تصفیه

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/522684>

