

عنوان مقاله:

تشخیص متون توهین آمیز در موتورهای جستجو با استفاده از یادگیری ماشین

محل انتشار:

هفتمین کنفرانس بین المللی وب پژوهی (سال: 1400)

تعداد صفحات اصل مقاله: 10

نویسندگان:

نیما سیفی - دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه گیلان، رشت، ایران

مهدی امینیان - استادیار، گروه مهندسی کامپیوتر، دانشگاه گیلان، رشت، ایران

خلاصه مقاله:

با توجه به گسترش محتوا در بسترهای رسانهای و ارتباطی مختلف و همچنین دسترسی کاربران به این امکانات، لزوم بررسی محتوای به اشتراک گذاشته شده به ویژه در ابعاد فرهنگی و اجتماعی به منظور ارائه داده های با کیفیت به افراد حاضر در این عرصه ها همواره احساس میشود. یکی از مسائلی که در محتوای متنی، به خصوص محتوای ویژه کودکان، فرهنگی، دانشگاهی و ... بسیار پر اهمیت است تشخیص متون توهین آمیز به کار برده شده است که در این مقاله به آن پرداخته میشود. با استفاده از یادگیری ماشین، (SVM) Naïve Bayes و (KNN) داده های پیش پردازش شده را به مدل مورد نظر آموزش میدهیم و انتظار داریم که خروجی مدلی باشد که با دریافت متن احتمال رکیک بودن محتوا را تشخیص دهد. داده های مورد نظر مجموعه ای از جستجو های انجام شده در یک موتور جستجوی فارسی هستند که به منظور افزایش محتوا، دوباره این عبارات را در گوگل جستجو کرده و صفحه اول نتیجه را به داده ها اضافه می کنیم. سپس تشخیص میدهیم که داده مورد نظر رکیک میباشد یا خیر (برچسب گذاری). مدل مورد نظر این داده ها را یادگیری کرده و پس از آن مدلی داریم که میتواند احتمال رکیک بودن داده ورودی را تشخیص دهد. نتایج بدستآمده نشان میدهد که معیار اندازه گیری صحت (Precision) در مدل های Naïve Bayes، SVM و KNN به ترتیب برابر با ۹۴.۰۵، ۹۷.۲۸ و ۸۶.۴۸ خواهد بود.

کلمات کلیدی:

یادگیری ماشین، پردازش زبان طبیعی، تشخیص کلمات رکیک، KNN، Naïve Bayes، SVM

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1236901>

