

دیتاست ارائه شده مجموعه داده مقالات انگلیسی که شامل فیلدهای مختلفی مانند عنوان مقاله، زمان انتشار آن، نویسندگان آن، متن مقاله، و غیره است. هدف از این پروژه ارائه مدل دسته‌بندی برای شناسایی مقالات مرتبط (با برچسب ۱ در ستون Relevant) است. برای این منظور از مدل‌های شناخته‌شده در کتابخانه Scikit-learn و مدل‌های pre-trained مانند BERT استفاده شود. همچنین unit of analysis های مختلف مد نظر قرار گیرند.

انتظار می‌رود در این مسئله مدل‌های دسته‌بندی مختلف پیاده‌سازی شوند و نتایج آنها با معیارهای ارزیابی مختلف با هم مقایسه شوند. خروجی مورد نظر گزارشی با رعایت اصول و استانداردهای نگارشی برای تدوین متون تخصصی است که شامل توضیح کامل مسئله، مجموعه داده، روش‌ها و تکنیک‌ها، نتایج و تحلیل نتایج می‌باشد.

انتظار می‌رود شما بتوانید در طی انجام مراحل این پروژه موضوعات ناآشنا و ناشناخته را یاد بگیرید و دانش خود را افزایش دهید. مهلت انجام این پروژه ۵ تا تاریخ ۳۰ تیر ۱۴۰۱ است و لازم به ذکر است که این مسئله جزو پروژه‌های فعلی شرکت نبوده و بخش کوچکی از پروژه‌های تحقیقاتی و کاری و آموزشی است که در گذشته انجام شده و به پایان رسیده است.