

به نام خدا
درس مبانی یادگیری عمیق
پروژه پایانی

استاد درس : دکتر مرضیه داوودآبادی
دستیاران : مرتضی حاجی آبادی، امیررضا ویشه
دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
نیمسال اول تحصیلی ۱۴۰۲ - ۱۴۰۳



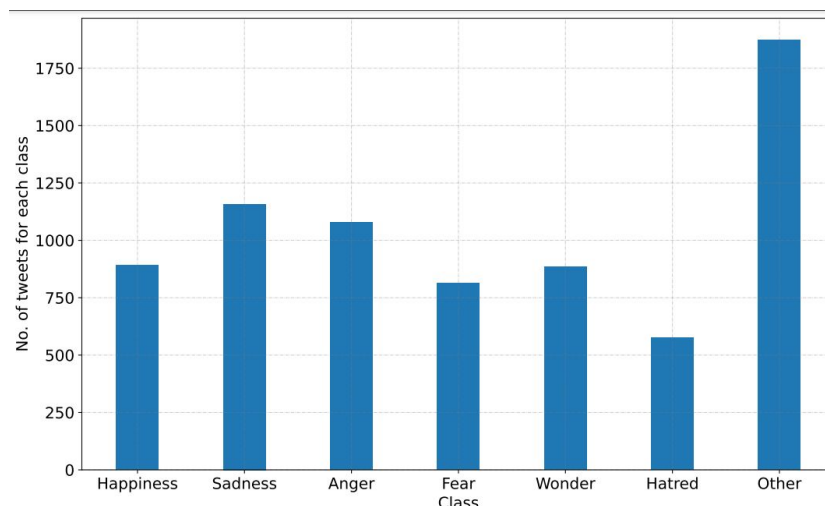
موضوع:

تحلیل احساسات در متن فارسی

احساسات و نظرات در متون آنلاین اهمیت زیادی دارند و توانایی دقیق در تشخیص احساسات در زبان فارسی یک کار چالشی است. در این پروژه، تلاش می‌شود تا با استفاده از مدل‌های یادگیری عمیق از پیش آموزش‌دیده، تحلیل احساسات متون فارسی صورت گیرد. در انجام این پروژه مجاز به استفاده از تمام ابزارهای خوانده شده در درس هستید. همچنین، در صورت استفاده از ابزارهایی که در درس مطرح نشده است لازم است به جزئیات آن مسلط باشید. جزئیات مختلف پروژه در ادامه آمده است که لطفاً با دقت مطالعه کنید.

۱ مجموعه داده

مجموعه دادگانی که برای این پروژه انتخاب شده است، مجموعه دادگان *ArmanEmo* است. این مجموعه دادگان برای تشخیص احساسات از متن فارسی است. این مجموعه داده شامل ۷ کلاس زیر است:



هدف اصلی شما در این پروژه آن است که بتوانید احساس برگرفته از یک جمله را در یکی از این ۷ دسته جای دهید. برای کسب اطلاعات بیشتر در مورد این مجموعه داده از این [مقاله](#) می توانید استفاده کنید.

۲ پیش پردازش داده ها

قبل از انتخاب مدل و آموزش آن، با تحقیق روی روش های پیش پردازش متون فارسی، پیش پردازش های لازم را روی داده ها انجام دهید.

۳ انتخاب مدل

به انتخاب خود و با در نظر گرفتن محدودیت های محاسباتی، یک مدل را برای آموزش روی مجموعه دادگان حاضر انتخاب کنید. محدودیتی در انتخاب مدل وجود ندارد. اولویت اول شما باید پیدا کردن مدل های مرتبط با زبان فارسی، به خصوص مدل هایی که برای اهداف مورد نظر ما مناسب هستند، باشد.

۴ یادگیری انتقالی

حال مدل انتخاب شده را روی مجموعه دادگان *ArmanEmo* برای بهبود توانایی تشخیص احساسات در متون فارسی تنظیم دقیق^۱ کنید.

^۱Fine tune

۵ ارزیابی مدل

مدل آموزش دیده را بر اساس معیارهای $recall$, $precision$ و $f1 - score$ ارزیابی کنید.

۶ بخش امتیازی

در این بخش، شما باید به دنبال یک مجموعه داده بگردید که هم از متن و هم از تصویر برای تشخیص احساسات استفاده می‌کند. پس از پیدا کردن مجموعه داده مناسب، شما باید مدل مناسبی را برای این داده‌ها پیدا کنید. پس از آموزش مدل روی این داده‌ها، شما باید تست‌های خود را روی چند داده انجام دهید و نتایج گزارش کنید. توجه داشته باشید، این بخش اختیاری است و فقط برای کسانی است که می‌خواهند چالش بیشتری را بپذیرند.

۷ خروجی

- یک گزارش از تمام کارهایی که در این پروژه انجام داده‌اید و نتایجی که بدست آورده‌اید. در نوشتن گزارش توجه داشته باشید که به هر مرجعی که استفاده می‌کنید (چه از لحاظ تئوری، چه از لحاظ کدنویسی) به دقت ارجاع دهید. یک قالب برای نوشتن گزارش در اختیار شما قرار خواهد گرفت.
- پیاده‌سازی مربوط به خواندن مجموعه داده‌گان، پیش‌پردازش داده‌ها، آموزش مدل و تابع $predict$ باید بارگذاری شوند. برای تابع $predict$ ایده‌آل است که این تابع دو حالت را پشتیبانی کند: یک حالت پایه که تنها یک توییت (متن) را پیش‌بینی کند و یک حالت دیگر که چند توییت یا متن را دریافت کرده و نتیجه کلی را ارائه دهد.
- نتایج کار شما از لحاظ کمی (دقت و سرعت) و کیفی با نتایج دیگران مقایسه خواهد شد.