

کافه بازار

cafebazaar.ir



به نام خدا

سلام،

امیدواریم حالتون خوب باشه. از اینکه بر روی این تسک وقت می‌ذارید ممنونیم. در راستای فرایند مصاحبه، در این مرحله تسکی براتون در نظر گرفتیم که در ادامه توضیح داده میشه.

داده ها:

در پوشه data، دو داده csv به نام‌های data_labeled و data_unlabeled قرار دارد. data_unlabeled شامل دو ستون به نام‌های id و text می‌باشد که ستون id معرف هر سطر و ستون text یکی از کامنت‌های بازار است. در data_labeled یک ستون اضافه به نام label نیز وجود دارد که مشخص می‌کند کامنت مورد نظر informative هست یا خیر. منظور از informative کامنت‌هایی هستند که حاوی اطلاعات مفیدی می‌باشند و صرفاً بیان نظرات شخصی و احساسات نسبت به یک برنامه نیستند. به طور کلی کامنت‌هایی informative محسوب می‌شوند که:

گزارشی از عملکرد اپ باشند:

- برنامه کرش می‌کند.
- بازی به مرحله‌ی بعد نمیره.
- پول ام رو پس نمیده .
- موقع نصب مشکل داره یا دسترسی عجیب می‌خواه.
- مناسب کودکان نیست یا هست.
- فلان مرحله یا بخش اپ این مشکل رو داره.

کامنت informative باید راجع به اپلیکیشن یا سرویسی که اپ ارائه می‌دهد باشد. کامنت راجع به توسعه دهنده informative محسوب نمی‌شود.

توضیحی از کارکرد اپ می‌دهد یا متن منتقد طوری ارائه می‌دهد:

- ویژگی‌های فلان را دارد
- این کار را بکند بهبود پیدا میکنه
- این بهبودهایی که داده خوبه

کامنت‌هایی که در گروه بالا نباشند و اطلاعات مهمی نداشته باشند هم Non-informative محسوب می‌شوند:

توصیف‌های کیفی:

- برنامه‌ی عالیست

- حتما نصب کنید
- سه ستاره کمه
- افتضاحه

کامنت‌هایی که به تشکر از محصول می‌پردازن informative محسوب نمیشن.

هدف:

هدف از این تسک، دسته‌بندی کامنت‌ها به دو دسته Informative و Non-informative است. در این راستا از شما انتظار می‌رود

کارهای زیر را انجام دهید:

- مراحل مرتبط با آشنایی اولیه با داده‌ها و انجام پیش‌پردازش‌های مناسب جهت تمیز کردن و آماده‌سازی برای استفاده جهت پیش‌بینی

لیبل هر کامنت

- آموزش مدل مناسب برای مسئله و بررسی و تحلیل نتایج

- اگر روش‌های دیگری برای بهبود دقت مدل میشناسید که به علت کمبود زمان یا موارد دیگر نتوانستید تست کنید به طور خلاصه

آن‌ها را معرفی کنید

دقت کنید که دو داده در اختیار شما قرار دارد، یکی با لیبل و دیگری بدون لیبل و انتخاب اینکه تنها از داده با لیبل استفاده کنید یا از هر دو

داده به عهده شماست و به روش انتخابی شما بستگی دارد.

لطفا سعی کنید تمام مراحل انجام شده به همراه کدها و در قالب یک گزارش برای ما ارسال کنید. اگر چند روش مختلف رو تست کردید نیاز

نیست روشی که منجر به بهترین جواب شده را ارسال کنید، سعی کنید تمام موارد رو ارسال کنید.

اگر سوالی پیرامون تسک داشتید، حتما بهمون ایمیل بزنید. ما تلاش می‌کنیم در سریع‌ترین زمان ممکن جواب ایمیلتون رو بدیم.

موفق باشید.