

فاز دو - خوشه بندی و طبقه بندی

هدف از این فاز آشنایی با انجام عملیات های خوشه بندی (Clustering) و طبقه بندی (Classification) بر روی دیتاستی که در اختیار دارید و بررسی نتایج آنها می باشد.

Embedding

در این بخش لازم است متن های clean شده ای که در فاز قبل به دست آورده بودید را با استفاده از BERT تبدیل به vector نموده تا برای ماشین قابل فهم بوده و بتوانید عملیات های خوشه بندی و طبقه بندی را بر روی آنها پیاده سازی نمایید. برای استفاده از BERT در این فاز از پروژه می توانید از مدل های پیش آموزش داده شده آن که در کتابخانه های مختلف مانند Hugging Face یا sentence-transformers و ... موجود است استفاده نمایید.

Clustering

در این بخش لازم است از خروجی عملیاتی که در بخش قبل انجام داده اید به عنوان ورودی الگوریتم های مختلف خوشه بندی استفاده نموده تا بتوانید دروس را خوشه بندی نمایید. هدف پیدا کردن دروسی است که شبیه هم باشند. لازم است پس از به دست آوردن خوشه های مربوطه آنها را بازبینی نموده و روش های ابتکاری و خلاقانه خود برای افزایش دقت بین خوشه های تشخیص داده شده را ارائه نمایید.

Classification (نمره اضافه)

برای این بخش لازم است شما یک مسئله طبقه بندی بر روی دیتاستی که در اختیار دارید تعریف و اقدام به حل آن نمایید. مثلاً شما می توانید مدلی آموزش دهید که گروه آموزشی درس را بر اساس اطلاعات درس پیش بینی میکند. لازم است پس از پیش بینی اقدام به محاسبه معیار های مختلف بررسی طبقه بند نمایید. (هدف رسیدن به مدلی با دقت بالا نیست و بکارگیری روش ها و الگوریتم های مختلف مورد توجه است.)

نکات تحویل

- لازم است با استفاده از کتابخانه matplotlib یا دیگر کتابخانه های مشابه، بصری سازی های لازم جهت درک خروجی هر مرحله از کد خود را انجام نمایید.
- این فاز در قالب گروه های ۲ نفره و یا به صورت انفرادی قابل انجام میباشد.
- در این فاز شما می بایست فایل notebook خود را بدون حذف خروجی های حاصل از اجرا ارسال نمایید.
- فایل ها باید در قالب Student1Name-Student2Name-CourseCode-phase2.zip ارسال شود.
- CourseCode برای گروه ساعت ۴ عدد ۱ و برای گروه ساعت ۶ عدد ۲ می باشد.
- ارسال فایل تنها از طریق سامانه VU مورد قبول بوده و فایل های ارسال شده در تلگرام و ... تصحیح نخواهند شد.

