10 در صد داده ها رو از ابتدا جدا کنید برای تست، مابقی برای train و validation

نتایج قسمت های مختلف رو بر اساس 10 در صد داده تست گزارش کنید

همانطور که همه مستحضر هستید حل هر مساله شبیه این دو فاز data representation داره و یک فاز مدل سازی، فاز های زیر جز اصلی پروژه هستند

ابتدا فاز data representation را با TF-IDF انجام دهید، مدل هایی که باید انجام دهید الف) logestic regression ب) decision treeج) ada boost

برای هر مدل confusion matrix را بدست آورده و راجع به تمامی معیارهای طبقه بندی که در هفته ششم بود بحث کنید

**فاز دوم**

برای data representation از لینک های زیر استفاده کنید : توجه باید مفهوم هر کدام را بفهمید و در گزارش خود قید کنید هر یک چگونه و با چه الگوریتمی data representation می کند و مزایا و معایبش چیست

https://code.google.com/archive/p/word2vec/

https://nlp.stanford.edu/projects/glove/ https://cogcomp.org/page/resource\_view/106

https://fasttext.cc/docs/en/english-vectors.html

برای مدل مدلهای زیر را آزمایش کنید اول) XGBoost دوم)LGBM سوم)SVM چهارم) LSTM

مطمئنا فاز دوم پیچیدگی های خود را دارد و انتظار نمی رود یک دانشجو همه را انجام دهد، به هر نسبتی که انجام دهید (خوب انجام دهید) مطمنا نمره خوبی خواهید گرفت.