**مواد مورد نیاز**

سلام و احترام به شما خوبان

بنا را در کارگاه تحلیل داده متنی آینده بر این گذاشته‌ایم که می‌خواهیم کمی دست‌ به کد شویم و با برخی تکنیک‌های اولیه در حوزه‌ی پردازش متن آشنا شویم.

برای اینکه بتوانیم در کنار هم یک تجربه خوب از کار تحلیل داده‌ متنی داشته باشیم؛ نیاز داریم که در یک سری از موارد همه با هم مشترک باشیم که عبارت‌اند از :

1. ابتدا توجه کنید که اجرای برخی از دستورات این فایل بعد از کپی کردن از روی فایل به دلیل تبدیل شدن فایل به pdf ممکن است با مشکل مواجه شوند در این صورت دستورات رو به طور مستقیم خودتان بنویسید.
2. دانلود کردن دیتاست: وارد این [لینک](https://phanous.ir/divar-posts-dataset/) شوید و دیتاست منتشر شده از آگهی‌های دیوار را دانلود کنید و در پوشه‌ی input-data قرار دهید.
3. پایتون ۳ داشته باشیم.
4. می‌خواهیم از virtual environmentای که توسط conda درست می‌شود استفاده کنیم.
   1. بنابراین لطفا همه Anacondaای که به سیستم‌مون بخوره و مخصوص python3 باشه را از [اینجا](https://www.anaconda.com/download) دانلود کنیم.
   2. با توجه به مشخصات سیستم‌تون [اینجا](https://docs.anaconda.com/anaconda/install/) راهنمایی‌های خوبی برای نصب Anaconda داره.
5. همچنین کدهایی که می‌زنیم در دفتر‌های Jupyter خواهند بود و از همان‌جا هم اجرا می‌شوند.
   1. بنابراین لطفا همه Jupyter را نصب کنیم.
   2. [اینجا](http://jupyter.org/install) و [اینجا](http://jupyter.readthedocs.io/en/latest/install.html) توضیح‌های مناسبی برای نصب‌ش دارند.
   3. اگر با نصب conda، پایتون آن را به عنوان پایتون default انتخاب کنید، ۲ خط زیر را در ترمینال خود اجرا کنید.
      1. python -m pip install --upgrade pip  
         python -m pip install jupyter
6. حال که ۲ گام بالا را با موفقیت pass کردیم، virtual environmentمان را با استفاده از conda می‌سازیم.
   1. در ابتدا فایل environment.yml را دریافت کنید.
   2. حال در ترمینال خود دستور زیر را وارد کنید:
      1. conda env create -f environment.yml
      2. توجه کنید که فایل environment.yml در مسیر کنونی ترمینال باشد یا آدرس آن را به صورت دقیق وارد کنید.
   3. با اجرای دستور گفته شده، virtual environmentمان آماده می‌شود.
   4. برای activate کردن آن دستور زیر را اجرا کنید:
      1. source activate divar-nlp-workshop
   5. همچنین برای deactivate کردن نیز از دستور زیر استفاده کنید.
      1. source deactivate
7. در گام بعدی باید virtual environmentمان را به Jupyter معرفی کنیم و آن را اصطلاحا به kernelهای jupyter اضافه کنیم. کافی است با یک دستور ساده این کار را انجام بدیم.
   1. ابتدا virtual environmentمان را activate می‌کنیم.
   2. سپس دستور زیر را اجرا می‌کنیم
      1. python -m ipykernel install --user --name divar-nlp-workshop --display-name "divar-nlp-kernel"
   3. با اجرا شدن دستور بالا،‌ virtual environmentمان به مجموعه kernelهای jupyter اضافه می‌شود. همانطور هم که مشخص هست نام‌ش را divar-nlp-kernel گذاشته‌ایم.
8. حال که گام بالا را هم با موفقیت pass کردید لطفا کار‌های زیر را با خیال راحت انجام دهید.
   1. در یک مسیر دلخواه‌تون دفتر Example را قرار دهید و دستور زیر را در ترمینال‌تون که در مسیر مشخص شده‌ قرار دارد،‌ اجرا کنید.
      1. jupyter notebook
   2. با اجرای دستور گفته شده، Jupyter شروع به کار و آدرس مکان اجرایی‌ش را چاپ می‌کند.
      1. در صورتی که مشکلی نباشد، آدرس معمولا به شکل زیر است.
      2. [http://localhost:8888/?token=3c52838751e67bc61ec729d8c0a252eee41ee4ad684ebcd6](http://localhost:8889/?token=3c52838751e67bc61ec729d8c0a252eee41ee4ad684ebcd6)
      3. یعنی در localhost و پورت ۸۸۸۸.
      4. همچنین بابت مسائل امنیتی این آدرس معمولا به همراه یک توکن نیز هست.
   3. دفتر Example.ipynb را به توسط api خود jupyter باز کنید.
   4. در بخش بالایی jupyter notebook باز شده قسمت kernel را فشار دهید، سپس change kernel را فشار دهید و از میان گزینه‌های موجود divar-nlp-kernel را برگزینید.
   5. اگر این kernel درست عمل کند باید در بخش بالایی سمت راست علامت Trusted را مشاهده کنید، در صورتی که علامت not-trusted را مشاهده کردید آن را بفشارید و در پنجره باز شده گزینه قرمز رنگ Trust را بزنید.
   6. در انتها بروی Cell ابتدایی رفته، و آن را با استفاده از shift+enter اجرا کنید.
   7. درصورتی که به هیچ erorrای نرسیدید شما موفق شده‌اید.
9. در گام پایانی هم برای دارا شدن FastText به توضیحات مربوط به building fasttext python module دراین [لینک](https://fasttext.cc/docs/en/support.html) توجه کنید و دستورات مربوطه را اجرا کنید. در نهایت برای اطمینان از درست نصب شدن fastext وارد نوت‌بوک با کرنل divar-nlp-kernel شده و با import fasttext از درست نصب شدن آن اطمینان حاصل کنید.