

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](#) 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 environment**‌ای که توسط **conda** درست می‌شود استفاده کنیم.
 - a. بنابراین لطفا همه **Anaconda**‌ای که به سیستم‌مون بخوره و مخصوص **python3** باشه را از [اینجا](#) دانلود کنیم.
 - b. با توجه به مشخصات سیستم‌تون [اینجا](#) راهنمایی‌های خوبی برای نصب **Anaconda** داره.
 5. همچنین کدهایی که می‌زنیم در دفترهای **Jupyter** خواهند بود و از همان‌جا هم اجرا می‌شوند.
 - a. بنابراین لطفا همه **Jupyter** را نصب کنیم.
 - b. [اینجا](#) و [اینجا](#) توضیح‌های مناسبی برای نصبش دارند.
 - c. اگر با نصب **conda**، پایتون آن را به عنوان پایتون **default** انتخاب کنید، ۲ خط زیر را در ترمینال خود اجرا کنید.

```
i. python -m pip install --upgrade pip
python -m pip install jupyter
```

6. حال که ۲ گام بالا را با موفقیت **pass** کردیم، **virtual environment**‌مان را با استفاده از **conda** می‌سازیم.

a. در ابتدا فایل **environment.yml** را دریافت کنید.

b. حال در ترمینال خود دستور زیر را وارد کنید:

i. `conda env create -f environment.yml`

ii. توجه کنید که فایل **environment.yml** در مسیر کنونی ترمینال باشد یا آدرس

آن را به صورت دقیق وارد کنید.

c. با اجرای دستور گفته شده، **virtual environment** مان آماده می‌شود.

d. برای **activate** کردن آن دستور زیر را اجرا کنید:

i. `source activate divar-nlp-workshop`

e. همچنین برای **deactivate** کردن نیز از دستور زیر استفاده کنید.

i. `source deactivate`

7. در گام بعدی باید **virtual environment** مان را به **Jupyter** معرفی کنیم و آن را اصطلاحاً

به **kernel** های **jupyter** اضافه کنیم. کافی است با یک دستور ساده این کار را انجام بدیم.

a. ابتدا **virtual environment** مان را **activate** می‌کنیم.

b. سپس دستور زیر را اجرا می‌کنیم

i. `python -m ipykernel install --user --name divar-nlp-workshop
--display-name "divar-nlp-kernel"`

c. با اجرا شدن دستور بالا، **virtual environment** مان به مجموعه **kernel** های

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 لطفاً کارهای زیر را با خیال راحت انجام دهید.

a. در یک مسیر دلخواهتون دفتر **Example** را قرار دهید و دستور زیر را در ترمینالتون که

در مسیر مشخص شده قرار دارد، اجرا کنید.

i. `jupyter notebook`

b. با اجرای دستور گفته شده، **Jupyter** شروع به کار و آدرس مکان اجرایش را چاپ

می‌کند.

i. در صورتی که مشکلی نباشد، آدرس معمولاً به شکل زیر است.

ii. <http://localhost:8888/?token=3c52838751e67bc61ec729d8c0a252eee41ee4ad684ebcd6>

iii. یعنی در **localhost** و پورت ۸۸۸۸.

iv. همچنین بابت مسائل امنیتی این آدرس معمولاً به همراه یک توکن نیز هست.

c. دفتر **Example.ipynb** را به توسط **api** خود **jupyter** باز کنید.

d. 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.

e. 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.

f. در انتها بروی **Cell** ابتدایی رفته، و آن را با استفاده از **shift+enter** اجرا کنید.

g. در صورتی که به هیچ **error** نرسیدید شما موفق شده‌اید.

9. در گام پایانی هم برای دارا شدن **FastText** به توضیحات مربوط به **building fasttext**

python module در این [لینک](#) توجه کنید و دستورات مربوطه را اجرا کنید. در نهایت برای

اطمینان از درست نصب شدن **fasttext** وارد نوت‌بوک با کرنل **divar-nlp-kernel** شده و با **import fasttext** از درست نصب شدن آن اطمینان حاصل کنید.