



Assignment 3: FastText

Description:

Requirements:

- Pick many documents → Yelp Dataset "yelp_academic_dataset_tip.json" (Use only Text column)
- Preprocessing to text

Step-1:

- Train FastText using genism on preprocessed texts → generate/save your own FastText model

Test your model:

- Print 10 similar words & 10 opposite words for a given/input words

Step-2:

Test pretrained model:

- Download pretrained FastText model <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.en.300.bin.gz>
- Print 10 similar words & 10 opposite words for a given/input words

Step-3:

- Update the pretrained model using your own data (Yelp_tip)
- Print 10 similar words & 10 opposite words for a given/input words

In a Pdf Report:

- Compare the output of the three steps for the same (given/input words)
- You must have a conclusion/comments

Delivery:

- The assignment will be done in **teams of three**.
- The assignment discussion will be held at lab time “After Eid” (the grades will be on discussion only there is no submission).