NLP Final Task

Introduction

Many countries speak Arabic; however, each country has its own dialect, the aim of this task is to build a model that predicts the dialect given the text.

Guidelines

- You are given a database having the needed data.
- You need to retrieve the data from different tables in the database and then start dealing with it.
- Target dialects are 5.
- The dataset and the dialect identification problem were addressed by Qatar Computing Research Institute, moreover, they published a paper, feel free to get more insights from it, https://arxiv.org/pdf/2005.06557.pdf
- You must use python.
- Choose the most suitable data pre-processing techniques.
- Train **two** models, a machine learning model and a deep learning model, then compare the results (You are free to choose any ML algorithm and any DL architecture)
- Use Flask or FastAPI or any suitable web framework to deploy the model locally.

Deliverables

1. Data Fetching script/notebook
2. Data Preprocessing script/notebook
3. Model Training script/notebook
4. Deployment script/notebook
5. Any additional files/documentation you need

Notes

- In the assessment phase, you'll be asked to run your models locally, furthermore, you'll be asked in any technical decision/implementation you've made, so **be well prepared**, and avoid over-complicated approaches you don't fully grasp.
- Early submission doesn't affect your grade, take your time.

NOTE: Make sure to prepare a powerpoint presentation summarizing your approach, data pre-processing, model architecture, evaluation metrics and results.

(Max 10 slides)