Natural Language Processing – Group Project – Installation Guide

Synthetic Datasets, Entity Recognition and Classification Models in NLP

Through this guide, we will explain how to set-up our Streamlit Application on Medical Diagnosis and Geo-Pattern detection.

Prerequisites

You will need Anaconda Navigator/Miniconda to run this app locally. If you do not have this application, you should download it first.

Installation Steps to run it locally.

1. Open a Terminal on Mac/Linux and Anaconda prompt on Windows.

2. Within the Terminal/Anaconda Prompt Terminal, create a python environment on version 3.10.14. We will name it streamlit_NLP.

```
conda create -n streamlit_nlp python=3.10.14
```

2. Once we have created the conda environment ready, we will activate it through the following command.

```
conda activate streamlit_nlp
```

3. We are now ready to install the streamlit module in our environment.

```
!pip install streamlit
```

4. After that, head to where you have the application folder and copy your path. Write in the terminal the following command to change the current directory to the Streamlit application.

```
cd /path_of_the_folder/NLPMedicalApp
```

5. Before running anything, we must write the following command to download the requirements to run it. These have been stored by us in the requirements.txt file.

```
pip install -r requirements.txt
```

6. Run this following command, a browser tab should open with address localhost:8501. You are all set! Enjoy our App!

```
streamlit run Entity_recognition.py
```

Other Notebooks:

We have also compiled a set of notebooks that do much more than the application itself. We will list them here:

Synthetic Datasets through API

- ChatGPT 3.5Turbo - OpenAI - PAID.ipynb
- Llama3 70B - Groq Cloud - FREE.ipynb

Synthetic Datasets through Local Transformers (Fine-tuning LLMs)

- Fine-Tuning and creating our own GPT2.ipynb

NLP Basic Classification Model (TFIDF LOGIT model development, available in App)

- NLP_Medical_Disease_Prediction_word2vec

NLP IR Medical Entity Recognition (Geo-trends based on Symptom Keywords, available in App)

- NLP_Medical_NER_Trasnformer_Geotrends

We have also deployed our App in Streamlit's server. You can check it in the URL down below:

https://medicalinsights.streamlit.app/