←Go Back to Introduction to Natural Language Processing

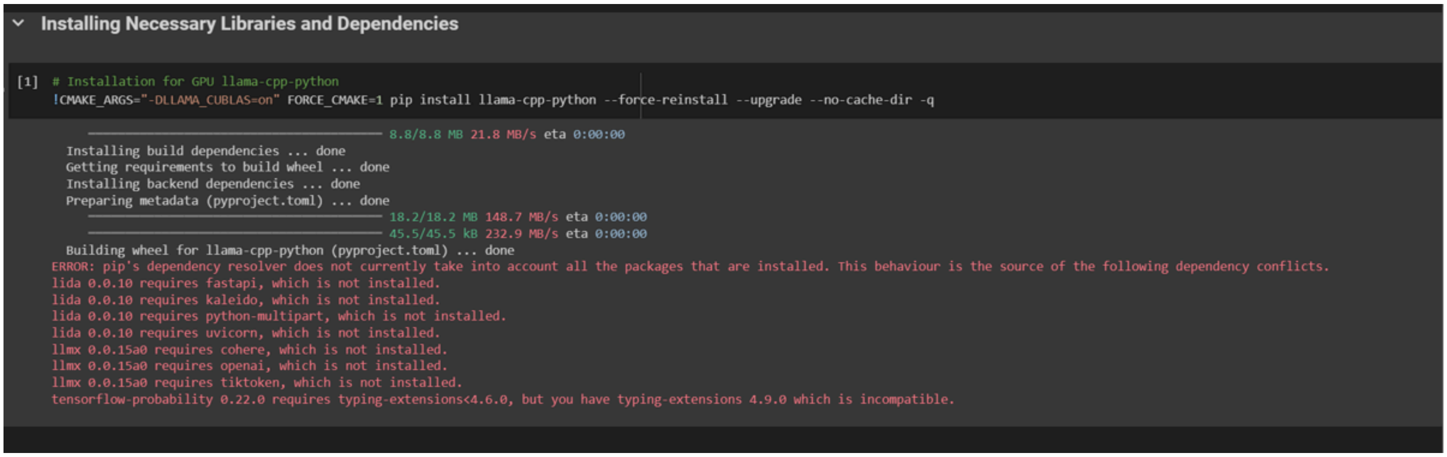
Course Content

FAQ - Stock Market News Sentiment Analysis and Summarization

1. How should one approach the project?

- Before starting the project, please read the problem statement carefully and go through the criteria and descriptions mentioned in the rubric.
- Once you understand the task, download the dataset and import it into a Python notebook to get started with the project.
  - Kindly use Google Colab for this project.
- To work on the project, you should start with a quick overview of the data
- Then, you can use the data to build a model.
- It is important to close the analysis with key findings and recommendations to the business.

2. I am getting the below error. How do I resolve it?



Installing Necessary Libraries and Dependencies

```
[1] # Installation for GPU llama-cpp-python
    !CMAKE_ARGS="-DLLAMA_CUBLAS=on" FORCE_CMAKE=1 pip install llama-cpp-python --force-reinstall --upgrade --no-cache-dir -q

     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 8.8/8.8 MB 21.8 MB/s eta 0:00:00
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
  Preparing metadata (pyproject.toml) ... done
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.2/18.2 MB 148.7 MB/s eta 0:00:00
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 45.5/45.5 kB 232.9 MB/s eta 0:00:00
  Building wheel for llama-cpp-python (pyproject.toml) ... done
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.
lida 0.0.10 requires fastapi, which is not installed.
lida 0.0.10 requires kaleido, which is not installed.
lida 0.0.10 requires python-multipart, which is not installed.
lida 0.0.10 requires uvicorn, which is not installed.
llmx 0.0.15a0 requires cohere, which is not installed.
llmx 0.0.15a0 requires openai, which is not installed.
llmx 0.0.15a0 requires tiktoken, which is not installed.
tensorflow-probability 0.22.0 requires typing-extensions<4.6.0, but you have typing-extensions 4.9.0 which is incompatible.
```

Please make sure that Google Colab is set to use the T4 GPU. Without a GPU, you will get CUDA errors. Also, the model might not work optimally even after it is loaded if GPU is not used.

Kindly note that once you run the code in the screenshot above, you will see the following output:

```
Building wheel for llama-cpp-python (pyproject.toml) ... done
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of
lida 0.0.10 requires fastapi, which is not installed.
lida 0.0.10 requires kaleido, which is not installed.
lida 0.0.10 requires python-multipart, which is not installed.
lida 0.0.10 requires uvicorn, which is not installed.
llmx 0.0.15a0 requires cohere, which is not installed.
llmx 0.0.15a0 requires openai, which is not installed.
llmx 0.0.15a0 requires tiktoken, which is not installed.
tensorflow-probability 0.22.0 requires typing-extensions<4.6.0, but you have typing-extensions 4.9.0 which is incompatible.
```

As long as the message

```
Building wheel for llama-cpp-python (pyproject.toml) ... done
```

is displayed, and the required library is installed correctly. The rest of the error messages can be ignored as they will not affect the code execution.

3. How to convert the file into HTML for submission?

The notebook can be downloaded to the system in .ipynb format and then converted to an HTML (.html) file using this link.

4. While creating the new column 'model_response,' the code is taking a long time to execute completely. Am I doing something wrong? Do I need to subscribe to Colab Pro?

The long execution time might be occurring because the Colab runtime is not set to *T4 GPU*. Please follow the below steps to change the runtime to *T4 GPU*:

- Click on "Runtime" in the menu bar
- Select "Change runtime type" from the dropdown menu
- In the "Hardware accelerator" section, choose "GPU"
- You may see multiple GPU options; choose "GPU" if you specifically want a T4 GPU
- After selecting the GPU option, click on the "Save" button

It is important to note the following:

1. Even after setting the runtime to *T4 GPU*, the code takes ~1 minutes per response.
2. A Colab Pro subscription is not required to complete this project.

5. I am getting an empty value in the "*model_response_parsed*" column and the following error is displayed:

```
Error parsing JSON: Expecting value: line 1 column 2 (char 1)
```

Could you please help me in fixing this?

LLMs can sometimes not respond the way we want them to, and we might need to be more specific regarding the output. In the case of the project, we want the output as a JSON. However, the model might end up sharing extra text in addition to the JSON, and that might hinder the JSON parses function from performing optimally.

One way to resolve this would be to specifically add instructions regarding output format. For example, one can mention the following line at the end of the prompt:

Do not include any other text in the output except the JSON.

If this fails, one can manually extract the JSON keys and values from the model response. (you'll get an idea of what to extract from the rows where the response is in JSON format).

Once all the JSON keys and values are extracted and added to a separate column, *pd.json_normalize()* can be used to convert the JSON to pandas DataFrame columns.

6. What is a JSON?

JSON stands for JavaScript Object Notation. It is a format in which data is represented as key-value pairs. Each key is a string enclosed in double quotes, followed by a colon, and then the corresponding value. The key-value pairs are separated by commas, and the entire data structure is enclosed within curly braces.

{ "key1": "value1", "key2": "value2", "key3": "value3" }

- `"key1"`, `"key2"`, and `"key3"` are the keys.
- `"value1"`, `"value2"`, and `"value3"` are the corresponding values associated with each key.

Example: Person = { "name": "Alice", "age": 25 }

‹ Previous

Next ›

Privacy Terms