Intro to Hugging Face and Python

Dirk Wulff

MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

UNI BASEL

OUR SOFTWARE STACK

+

+

OUR SOFTWARE STACK

+

+

WHY IS HUGGING FACE?

WHY IS HUGGING FACE?

Traditional language modelling pipeline:

WHY IS HUGGING FACE?

Traditional language modelling pipeline:

1. Find out the model architecture

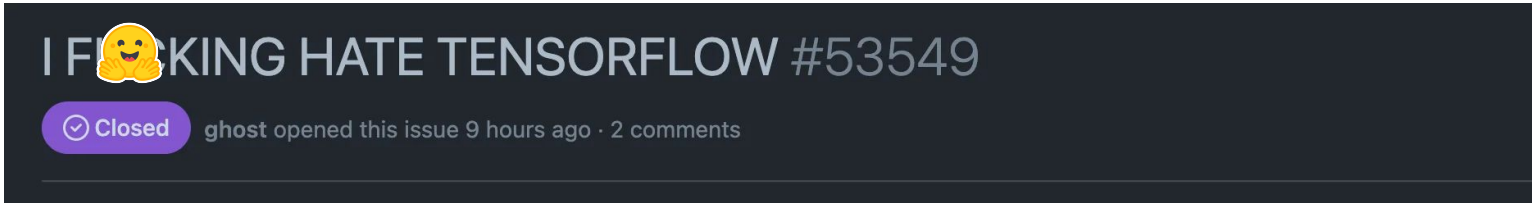
WHY IS HUGGING FACE? 

Traditional language modelling pipeline:

1. Find out the model architecture
2. Implement the model architecture in code with deep learning frameworks (e.g PyTorch/Tensorflow).

1. DEEP LEARNING LIBRARIES CAN BE DIFFICULT

1. DEEP LEARNING LIBRARIES CAN BE DIFFICULT



WHY HUGGING FACE?

Traditional language modelling pipeline:

1. Find out the model architecture
2. Implement the model architecture in code with deep learning libraries (e.g PyTorch/Tensorflow).

WHY HUGGING FACE?

Traditional language modelling pipeline:

1. Find out the model architecture
2. Implement the model architecture in code with deep learning libraries (e.g PyTorch/Tensorflow).
3. Load the pretrained weights (if available) from a server.
4. Process the inputs (using the correct tokenizer for the model)
5. Implement data loaders
6. Define a loss function
7. Stick a task-specific “head” on the model

HUGGING FACE PIPELINES

Import pipeline

```
from transformers import pipeline
```

Initialise pipeline

```
pipe = pipeline('text-generation', model='gpt2')
```

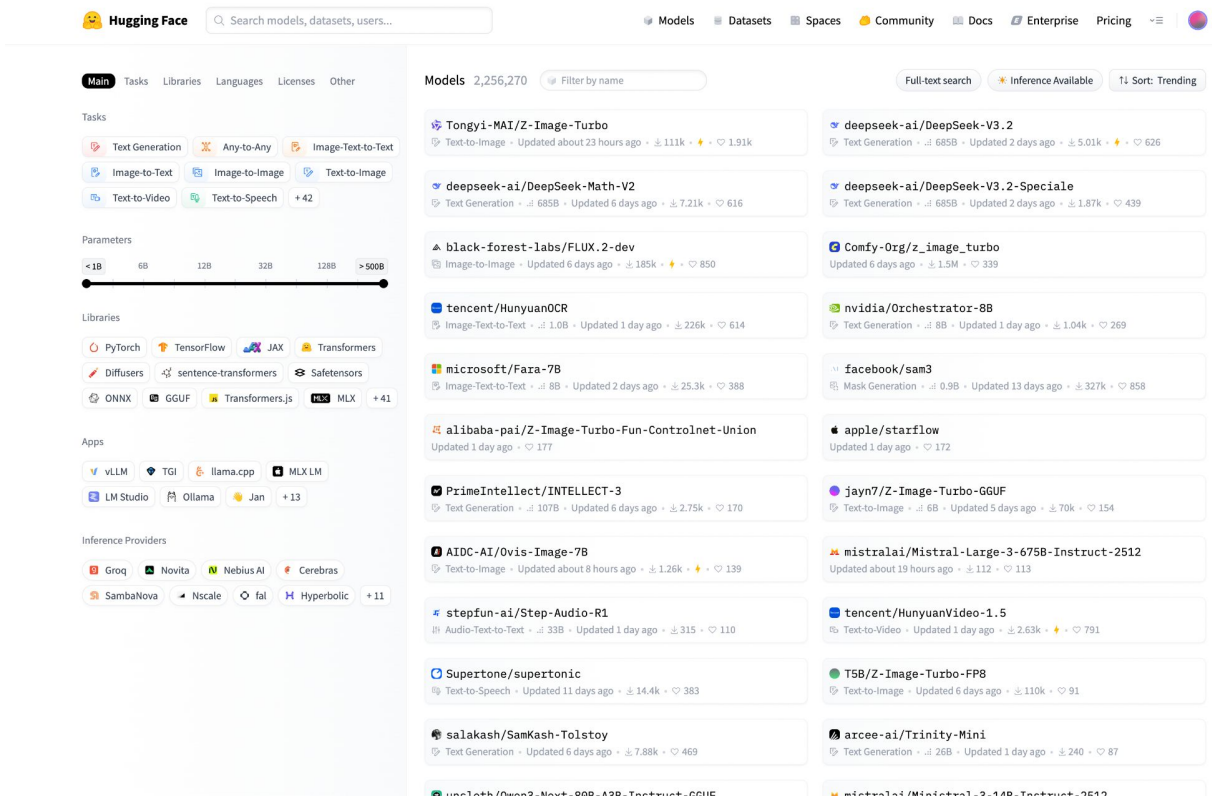
Load model input

```
prompt = """
Once upon a time in a land far far away, there was a young prince named
John. He was known for his bravery and courage. One day, he decided to go on
an adventure to explore the unknown lands.
"""`
```

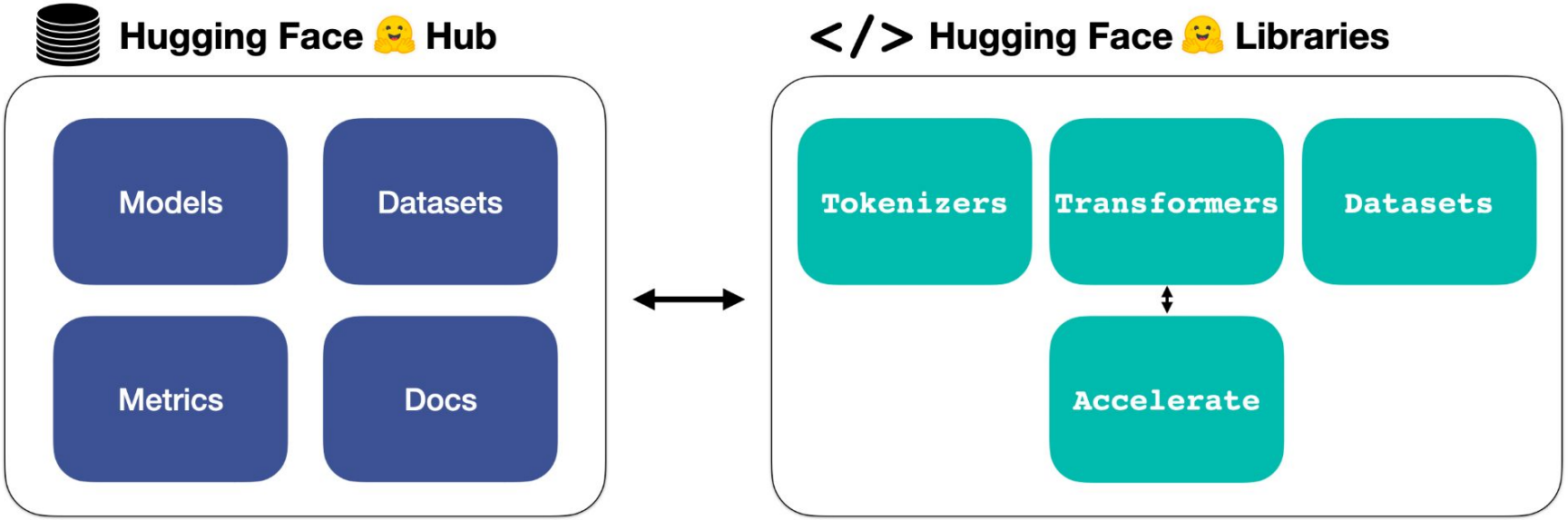
Feed input the model

```
output = pipe(prompt, max_length=100)
print(output)
```

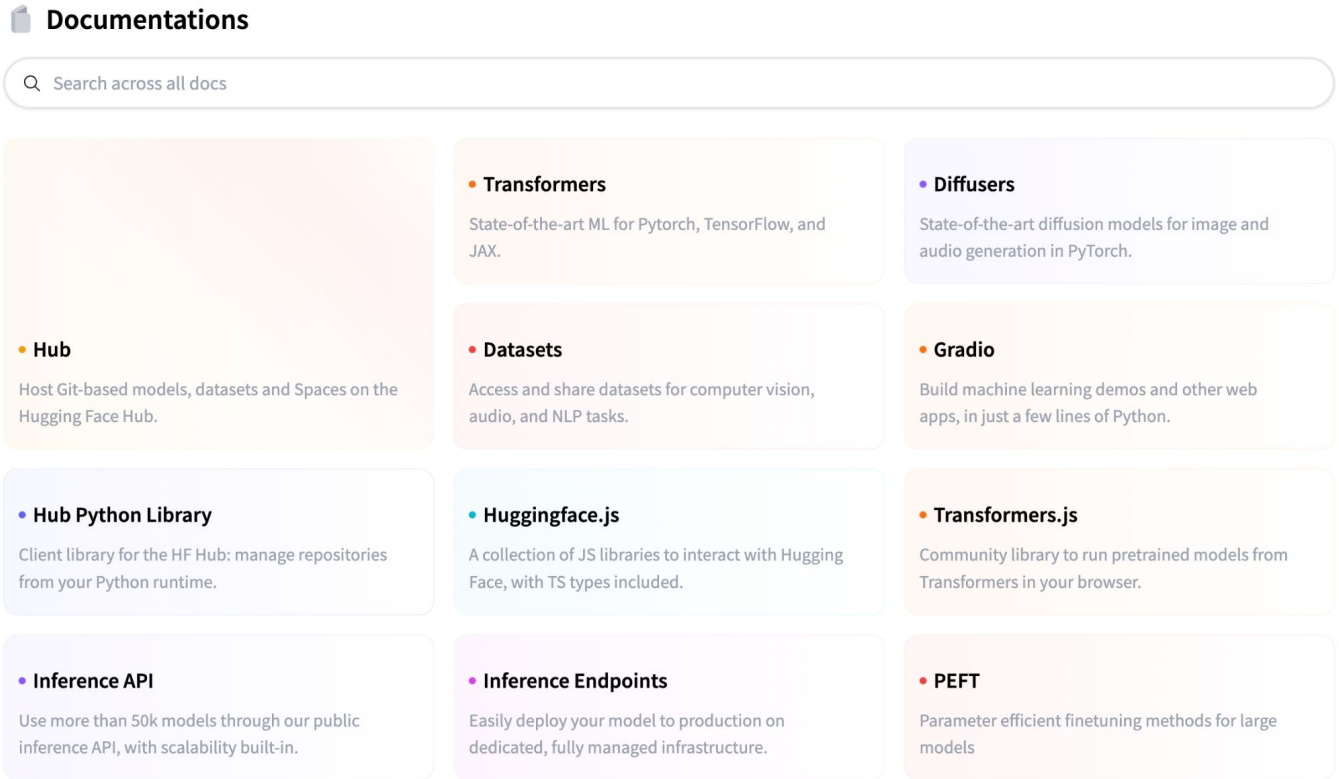
HUGGING FACE



THE HUGGING FACE ECOSYSTEM



HUGGING FACE DOCUMENTATION



OUR SOFTWARE STACK

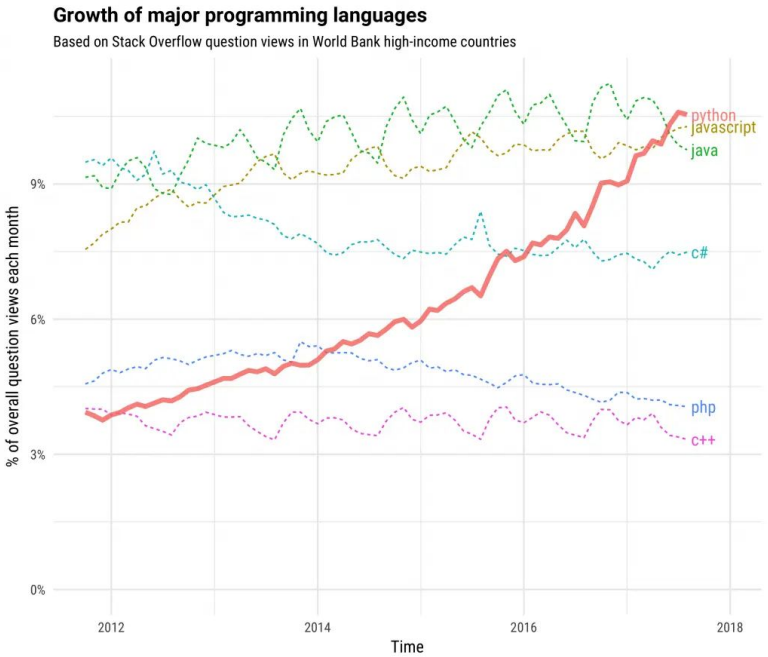
+

+

PYTHON 

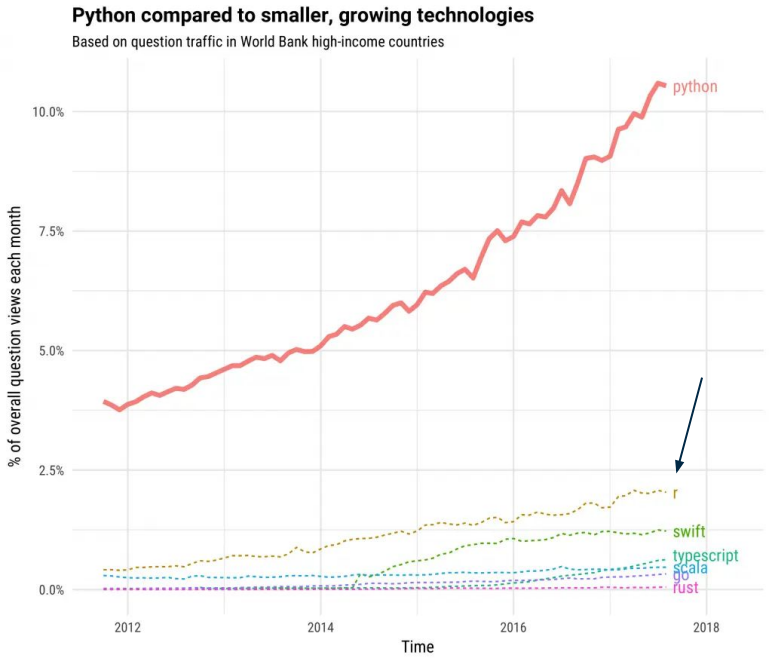
PYTHON

- A **high-level** programming language known for its simplicity and readability.
- Used in various domains such as web development, data analysis, **artificial intelligence**, and scientific computing.



PYTHON

- A **high-level** programming language known for its simplicity and readability.
- Used in various domains such as web development, data analysis, **artificial intelligence**, and scientific computing.



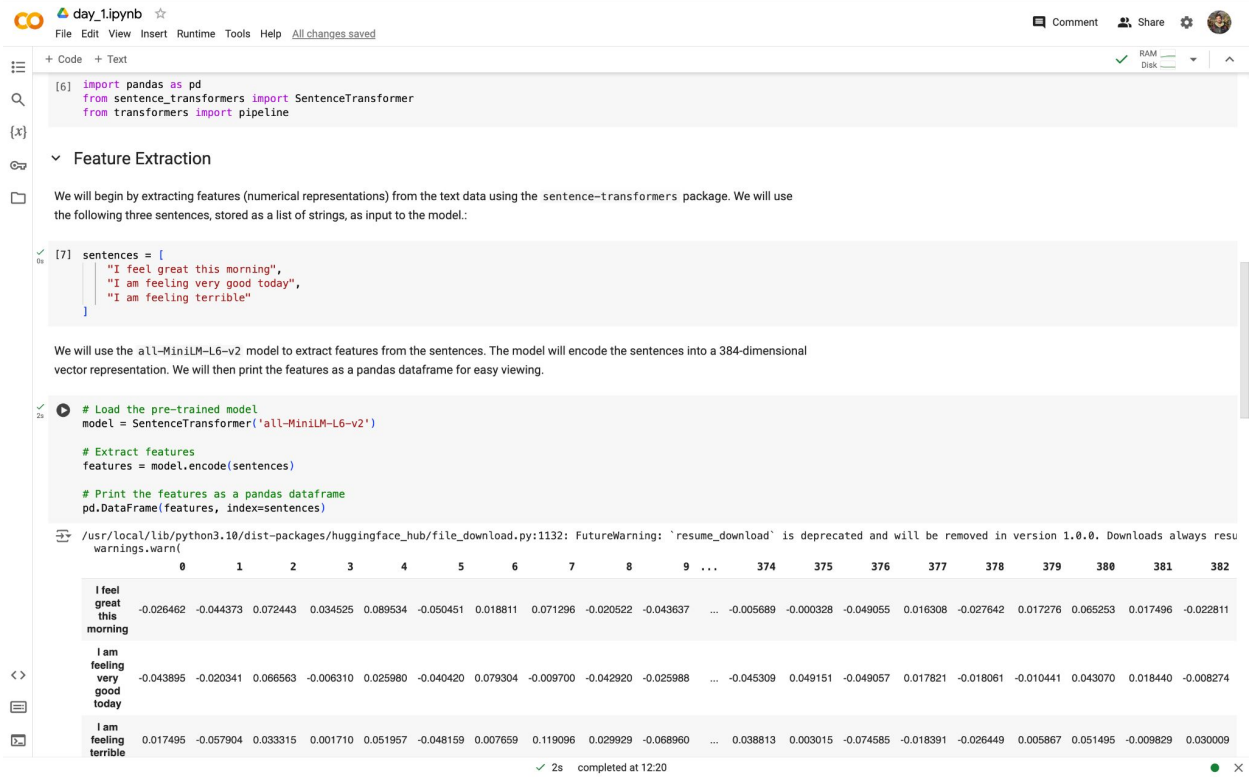
OUR SOFTWARE STACK

+

+

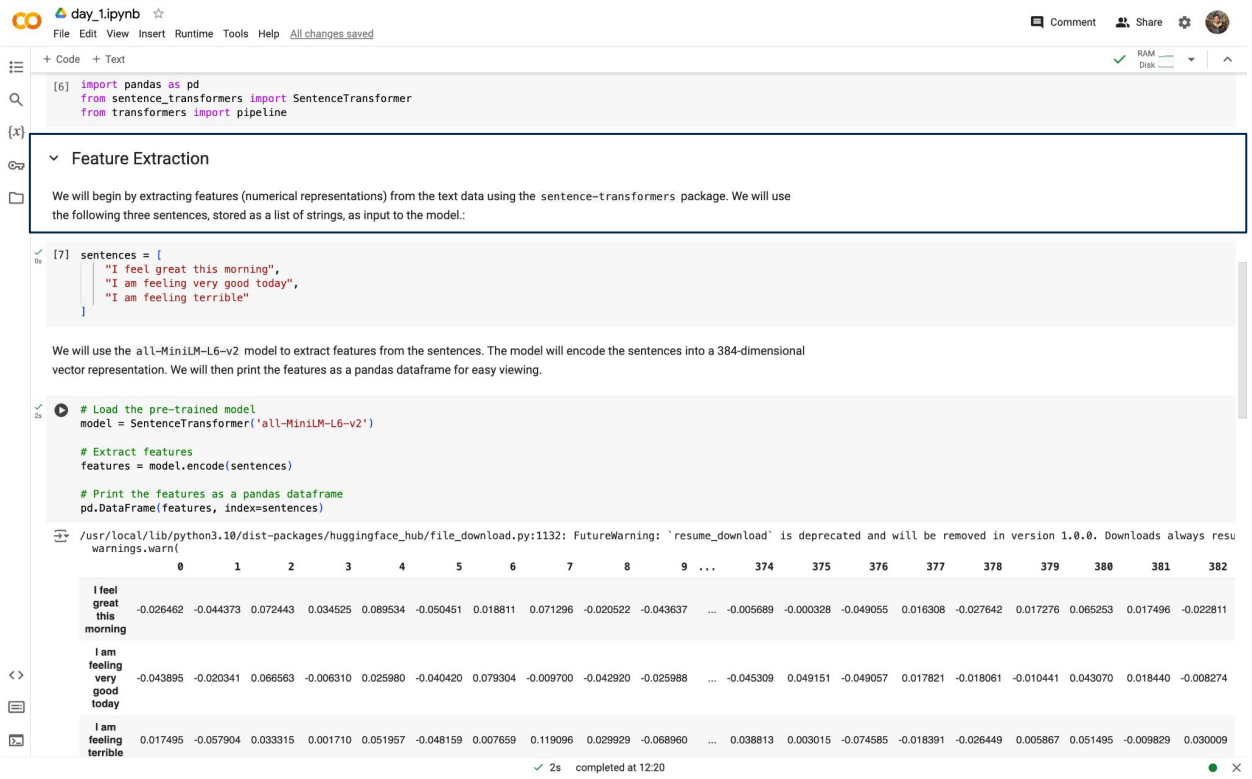
PYTHON  + Google Colab 

PYTHON + Google Colab



PYTHON + Google Colab

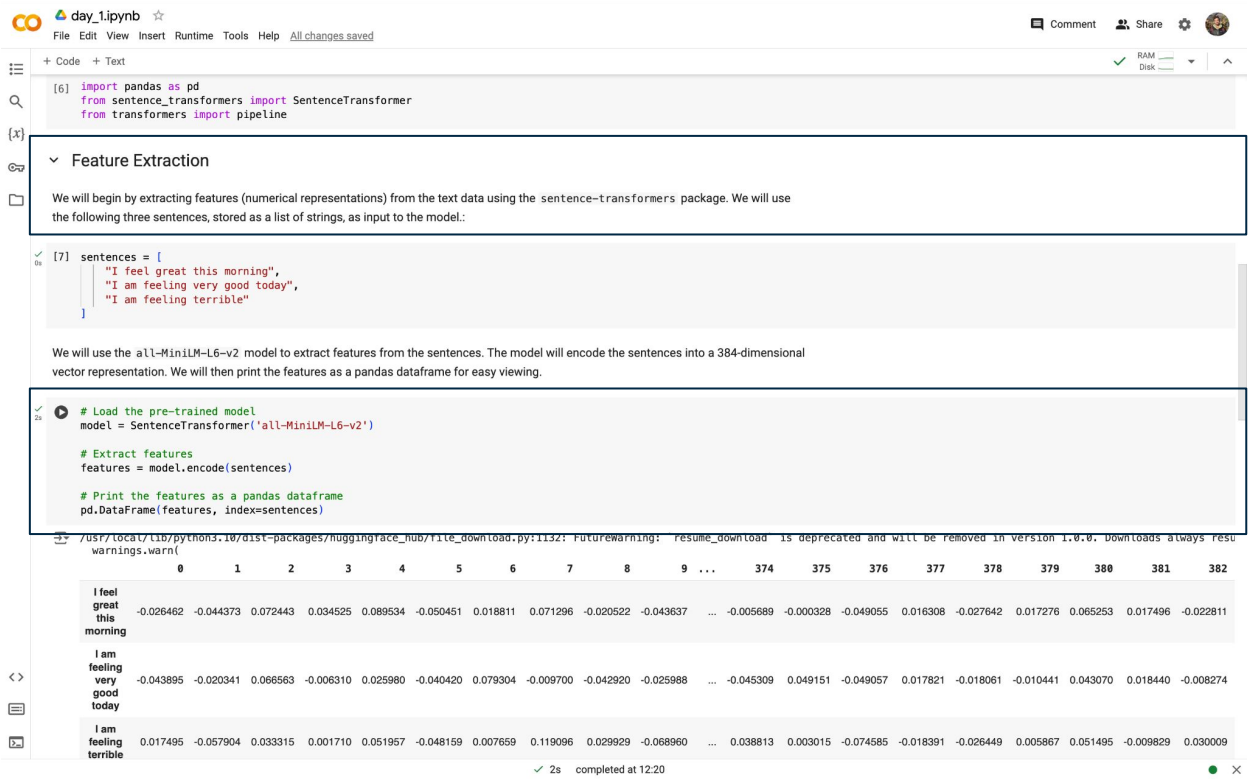
1. Markdown



PYTHON + Google Colab

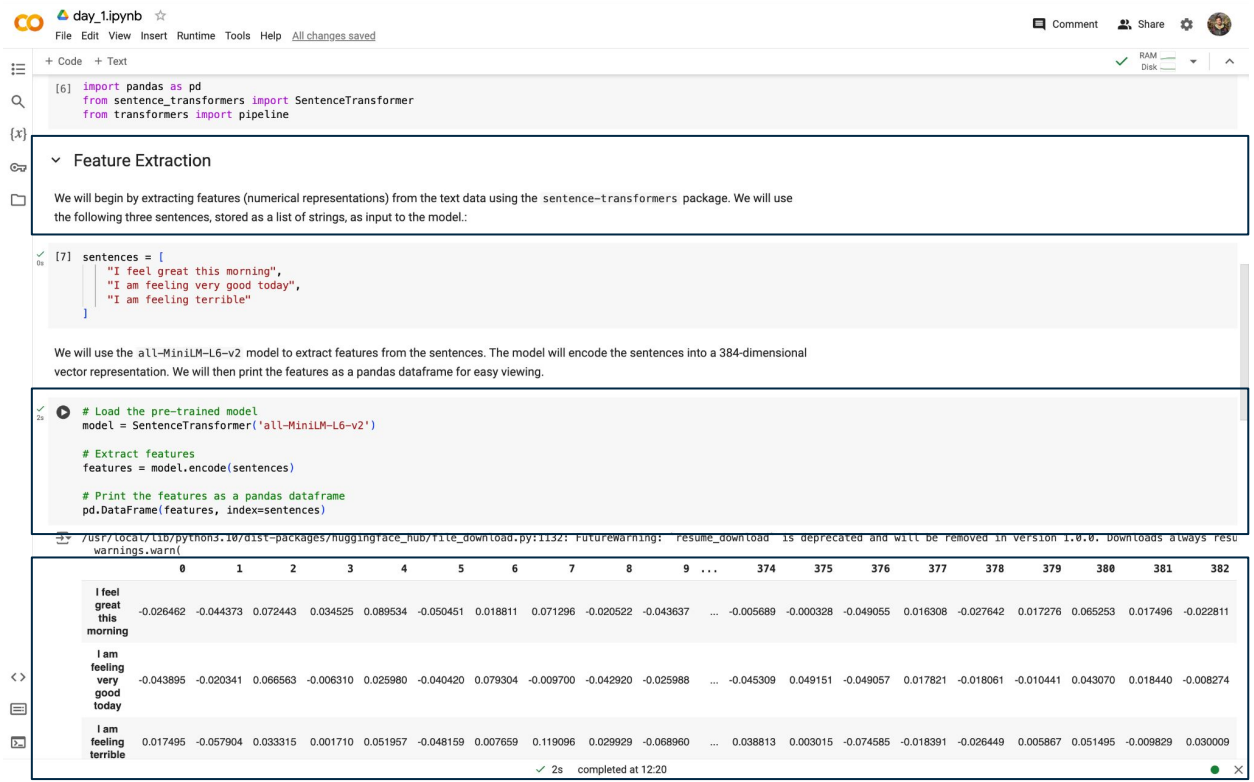
1. Markdown

2. Code

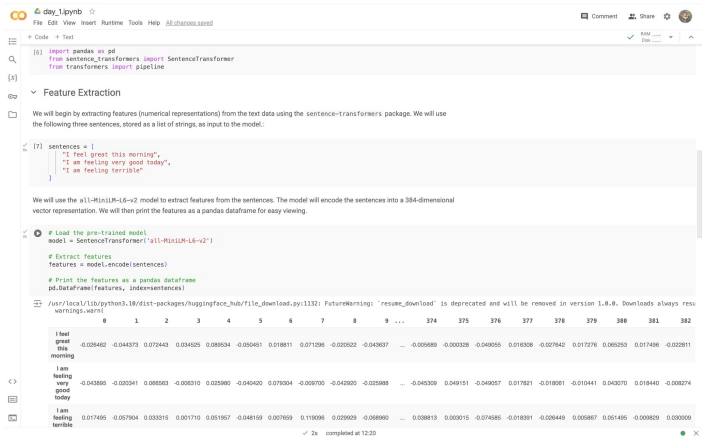


PYTHON + Google Colab

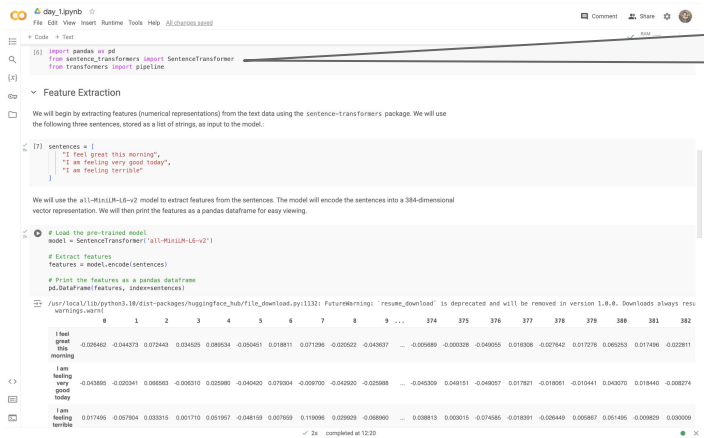
1. Markdown
2. Code
3. Printouts



PYTHON + Google Colab



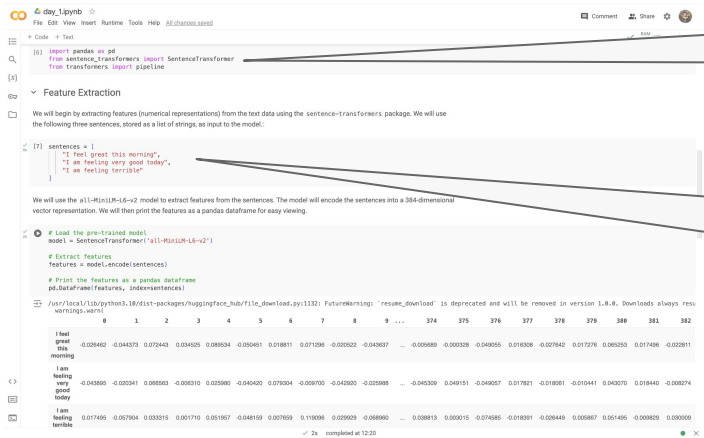
PYTHON + Google Colab



package imports

```
import pandas as pd
from sentence_transformers import SentenceTransformer
from transformers import pipeline
```

PYTHON + Google Colab



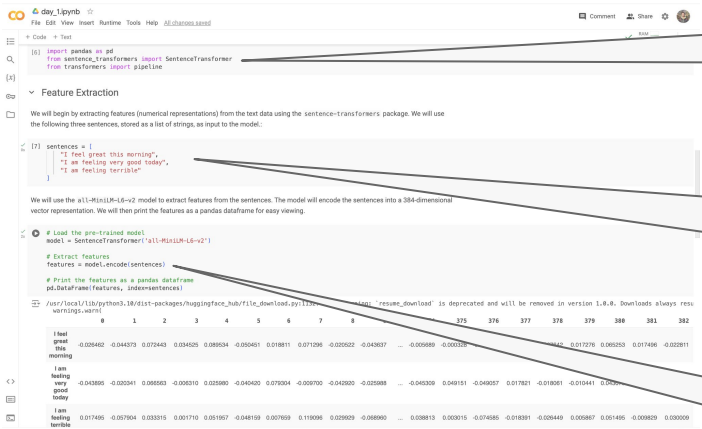
package imports

```
import pandas as pd
from sentence_transformers import SentenceTransformer
from transformers import pipeline
```

variable assignment, lists, strings

```
sentences = [
    "I feel great this morning",
    "I am feeling very good today",
    "I am feeling terrible"
]
```

PYTHON + Google Colab



package imports

```
import pandas as pd
from sentence_transformers import SentenceTransformer
from transformers import pipeline
```

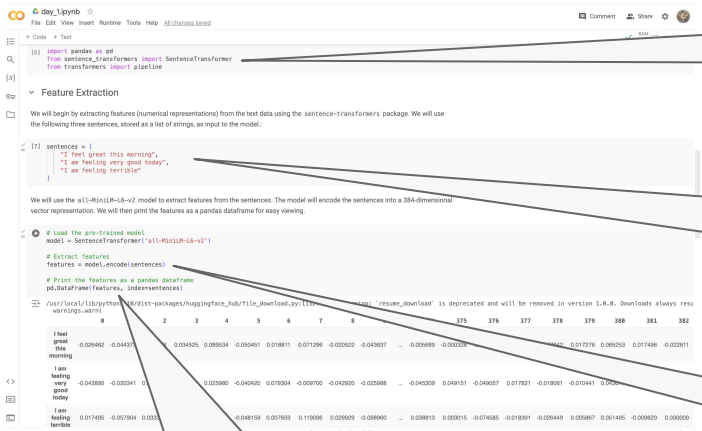
variable assignment, lists, strings

```
sentences = [
    "I feel great this morning",
    "I am feeling very good today",
    "I am feeling terrible"
]
```

dot notation, methods, attributes

```
# Extract features
features = model.encode(sentences)
```

PYTHON + Google Colab



package imports

```
import pandas as pd
from sentence_transformers import SentenceTransformer
from transformers import pipeline
```

variable assignment, lists, strings

```
sentences = [
    "I feel great this morning",
    "I am feeling very good today",
    "I am feeling terrible"
]
```

dot notation, methods, attributes

```
# Extract features
features = model.encode(sentences)
```

printing

```
# Print the features as a pandas dataframe
pd.DataFrame(features, index=sentences)
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/file_download.py:
  warnings.warn(
```

	0	1	2	3	4	5	6
I feel great this morning	-0.026462	-0.044373	0.072443	0.034525	0.089534	-0.050451	0.018811	0.
I am feeling very good today	-0.043895	-0.020341	0.066563	-0.006310	0.025980	-0.040420	0.079304	-0.

OUR SOFTWARE STACK

+

+

Time to install our stack…

MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

UNI
BASEL

LLM4BeSci_2025MetaRep

Public

Pin Watch 0 Fork 0 Star 0

main 1 Branch 0 Tags Go to file t Add file Code

dwulff Fix link formatting for Taisiia Tikhomirova 72cb358 · 7 minutes ago 12 Commits

day_1	init	yesterday
day_2	..	13 minutes ago
day_3	..	13 minutes ago
day_4	init	yesterday
day_5	init	yesterday
.DS_Store	..	13 minutes ago
LICENSE.txt	init	yesterday
README.md	Fix link formatting for Taisiia Tikhomirova	7 minutes ago
cover_metarep.png	Add files via upload	43 minutes ago
notes.txt	init	yesterday

README License

LLM4BeSci at MetaRep, MPIB Berlin, Dec 2025

About

No description, website, or topics provided.

Readme
View license
Activity
0 stars
0 watching
0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

Jupyter Notebook 100.0%