

Hugging Face i hosting projektów NLP

Grzyb Gerard
Jamróż Mateusz

Abstrakt

Projekt polega na przedstawieniu szybkiej możliwości znajdowania zbioru danych, wykorzystywanych istniejących modeli, oraz szybki i darmowy hosting oferowany przez platformę Hugging Face oraz możliwość hostingu przez Google Cloud Platform, oraz Amazon Web Services.

Wstęp

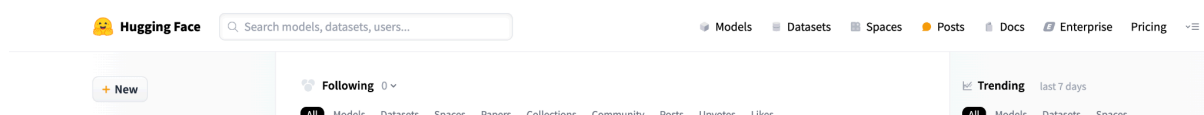
Celem projektu jest przedstawienie oferowanych możliwości platformy Hugging Face, wykorzystania dostępnych usług, oraz prezentacja możliwości hostingu modelu uczenia maszynowego.

W projekcie został poruszony zakres obsługi, budowania, oraz implementacji modeli uczenia maszynowego w środowiskach produkcyjnych przy użyciu dedykowanej platformy. Zbadano również zagadnienia takie jak konteneryzacja, monitorowanie modeli, wzorzec "gradio", transformersy, oraz "spaces".

Do celów badawczych zostały wykorzystane technologie takie jak: docker, python, "Google Cloud Platform", "Amazon Web Services", oraz platforma Hugging Face.

Opis

Głównym tematem opisywanego tematu jest platforma the Hugging Face. Jest platforma open-source do przetwarzania języka naturalnego (NLP) i sztucznej inteligencji, która pozwala na udostępnianiu projektów i modeli sztucznej inteligencji, oraz wspólną kolaborację.



Jednym z koniecznych kroków podczas nauki modelu jest pozyskanie zbiorów danych, które ta platforma oferuje. Dzięki bogatemu zbiorowi danych (w ilości ponad 250 tysięcy zbiorów) platforma pozwala na dokładne i precyzyjne zastosowanie ich do własnych potrzeb.

Datasets 250,885

HuggingFaceTB/smoltalk <small>Viewer · Updated about 2 hours ago · 2.2M · 1.04k · 133</small>	microsoft/orca-agentinstruct-1M-v1 <small>Viewer · Updated 26 days ago · 1.05M · 3.2k · 360</small>
fka/awesome-chatgpt-prompts <small>Viewer · Updated Sep 3 · 170 · 9.96k · 6.36k</small>	allenai/tulu-3-sft-mixture <small>Viewer · Updated 5 days ago · 939k · 923 · 39</small>
mLabonne/orca-agentinstruct-1M-v1-cleaned <small>Viewer · Updated 8 days ago · 1.05M · 988 · 45</small>	AI-MO/NuminaMath-CoT <small>Viewer · Updated 1 day ago · 860k · 2.44k · 238</small>
VTSNLP/vietnamese_curated_dataset <small>Viewer · Updated 2 days ago · 12.2M · 368 · 29</small>	PleIAs/common_corpus <small>Viewer · Updated 4 days ago · 397M · 49.1k · 164</small>
VTSNLP/instruct_general_dataset <small>Viewer · Updated Sep 30 · 4.53M · 264 · 25</small>	01-OPEN/Open01-SFT <small>Preview · Updated 4 days ago · 168 · 20</small>
microsoft/BiomedParseData <small>Preview · Updated 8 days ago · 168 · 20</small>	laion/LAIION-DISCO-12M <small>Viewer · Updated 12 days ago · 12.3M · 390 · 17</small>

Biblioteka Transformers oferuje narzędzia do łatwego pobierania i trenowania najnowocześniejszych modeli wstępnie wytrenowanych. Wykorzystanie takich modeli pozwala zmniejszyć koszty obliczeniowe, ślad węglowy oraz oszczędza czas i zasoby potrzebne na trenowanie od podstaw. Obsługiwane są zadania związane z: przetwarzaniem języka naturalnego, wizją komputerową, audio i multimodalnością.

How to use from the 🤗/transformers library

```
from transformers import AutoTokenizer, AutoModelForCausalLM

tokenizer = AutoTokenizer.from_pretrained("distilgpt2")

model = AutoModelForCausalLM.from_pretrained("distilgpt2")
```

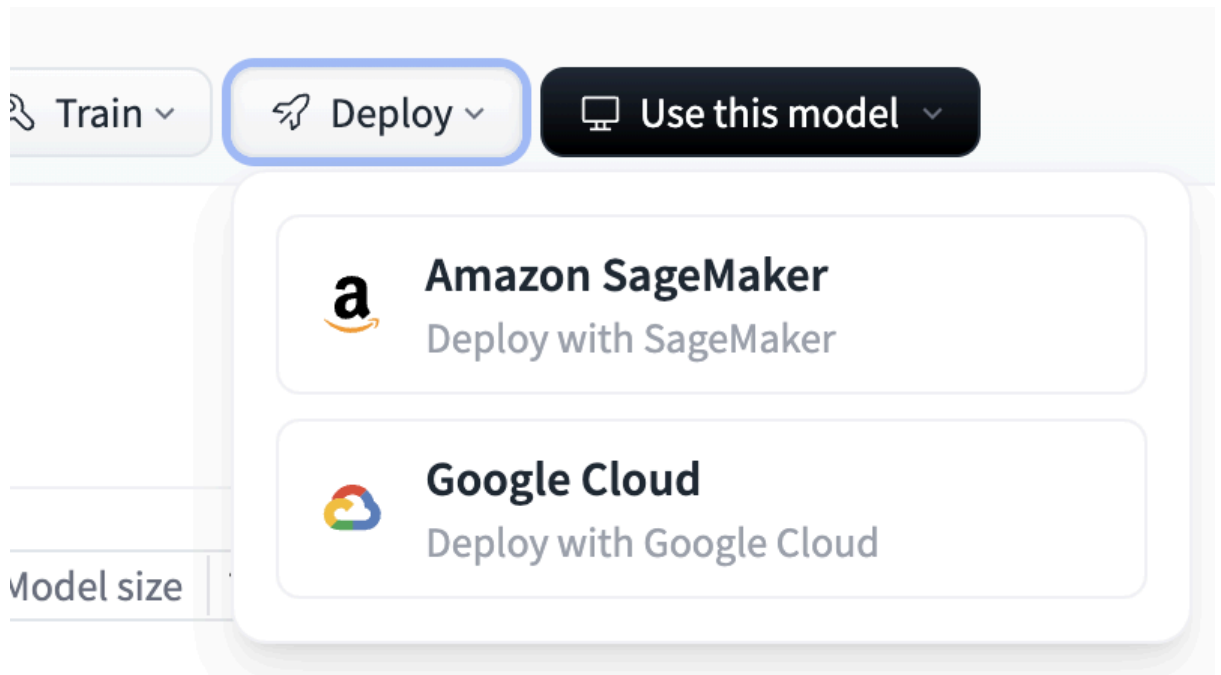
Or just clone the model repo

```
git lfs install
git clone https://huggingface.co/distilgpt2

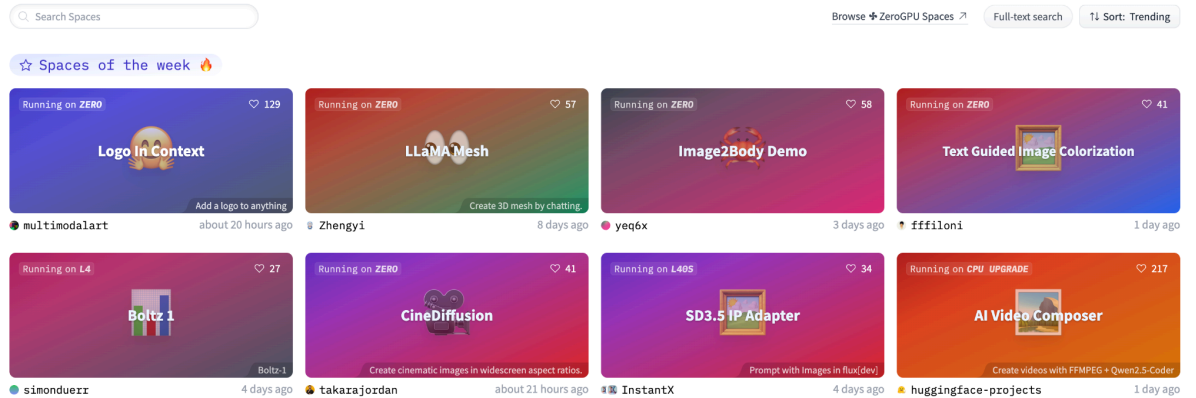
# if you want to clone without large files - just their pointers
# prepend your git clone with the following env var:
GIT_LFS_SKIP_SMUDGE=1
```

Następnym omawianym aspektem jest hosting, oraz wdrażanie do systemów produkcyjnych modeli uczenia maszynowego. Wyżej wspomniana platforma oferuje wiele możliwości, w tym:

1. Amazon SageMaker i Google Cloud → są to dedykowane rozwiązania od Amazon, oraz Google, dzięki którym proces wdrażania jest bezpieczniejszy, oraz dedykowany do korporacyjnych rozwiązań, lecz wiąże się to z większym progiem wejścia, oraz zaawansowaną konfiguracją



2. Spaces → Możliwość hostowania aplikacji ML na prostych serwerach wykorzystujących technologie takie jak Gradio czy Streamlit.



Platforma oferuje możliwość przeglądania istniejących wdrożonych modeli, ale również w prosty sposób wdrożenia własnego modelu.



Create a new Space

Spaces are Git repositories that host application code for Machine Learning demos. You can build Spaces with Python libraries like Streamlit or Gradio, or using Docker images.

Owner

matjarm



Space name

New Space name

Short description

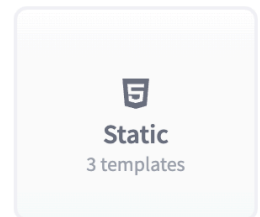
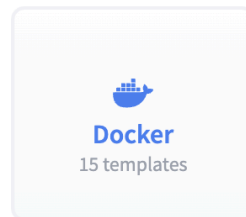
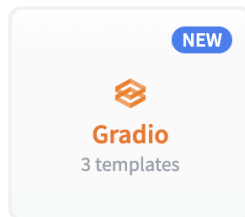
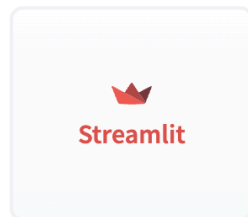
Short Description

License

License

Select the Space SDK

You can choose between Streamlit, Gradio and Static for your Space. Or pick Docker to host any other app.



Space hardware Free

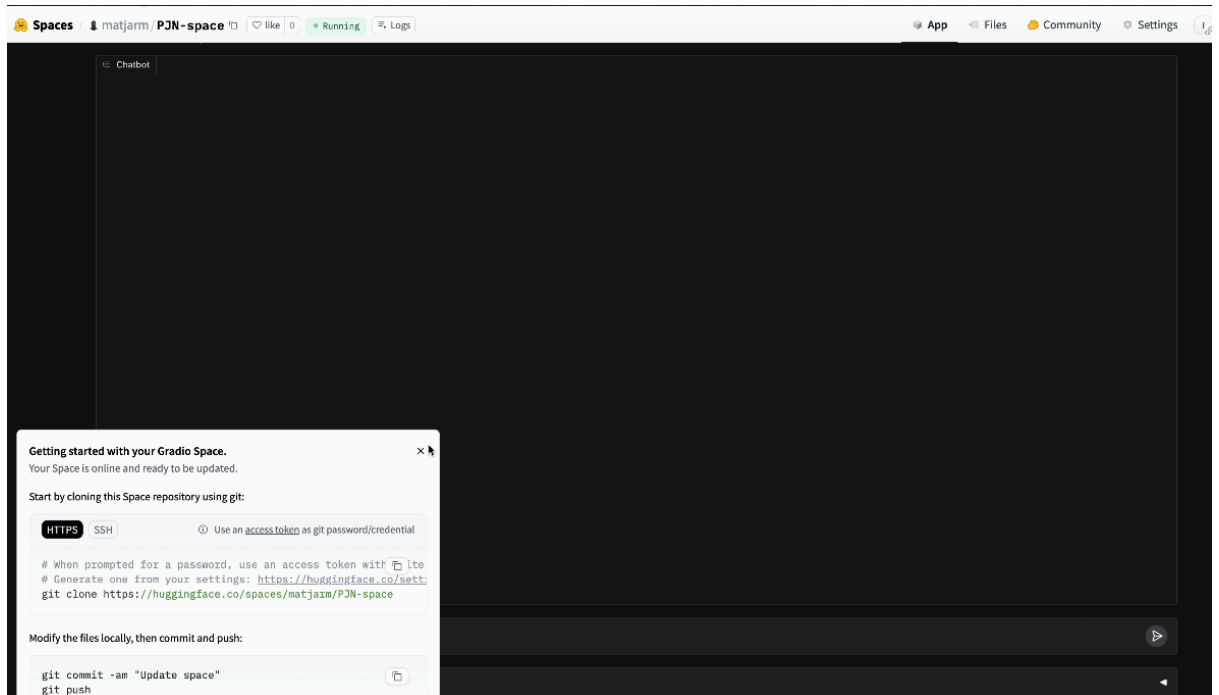
CPU basic · 2 vCPU · 16 GB · FREE



You can switch to a different hardware at any time in your Space settings.
You will be billed for every minute of uptime on a paid hardware.

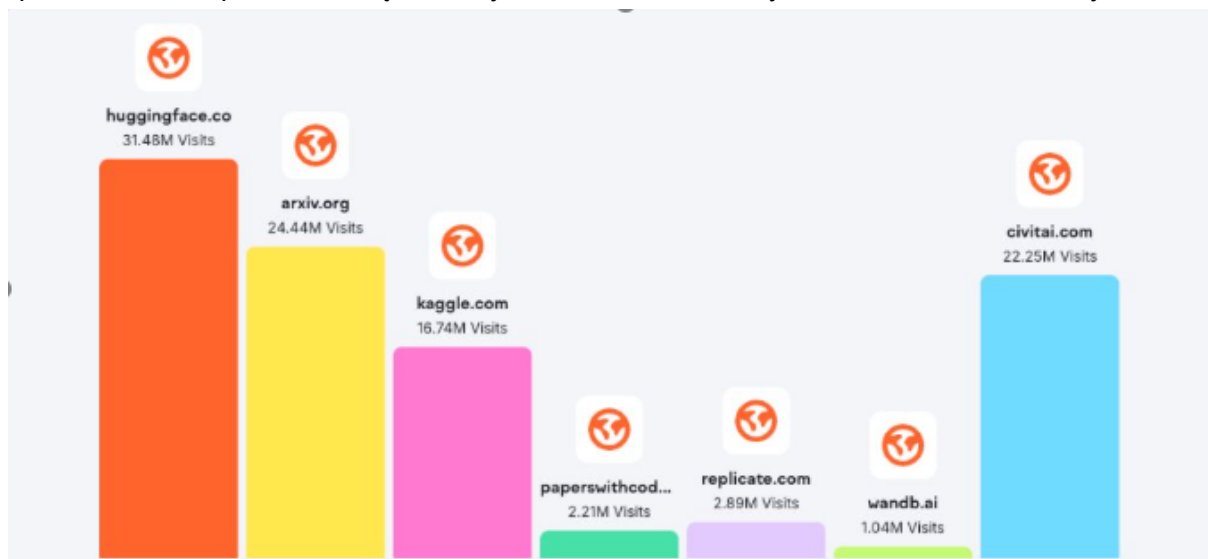
Poprzez wypełnienie formularza, oraz dostarczenie wzorca aplikacja została wdrożona w parę minut. Oferuje to dużą możliwość szybkiej modyfikacji, oraz rozszerzalności, gdyż wystarczy dostarczyć skrypt, lub model napisany w języku python.

Na poniższym zdjęciu zaprezentowane jest wdrożenie istniejącego chatbot'a, którego utworzenie nie wymaga szaty graficznej, oraz konfiguracji zdalnego hosta.



Kolejnym aspektem jest porównanie popularności innych platform oferujących podobne możliwości. Wspomniane platformy to między innymi Arxiv.org, kaggle.com czy civital.com.

Hugging Face jest najpopularniejszą platformą wśród konkurentów - ma największą społeczność, co przekłada się na duży zasób aktualizowanych modeli i zbiorów danych.



Podsumowanie

Hugging Face to narzędzie o ogromnym potencjale, zwłaszcza dla szybkiego prototypowania i współdzielenia modeli. Dzięki bogatemu zbiorowi danych i implementacji istniejących modeli w szacie graficznej ułatwia to proces implementacji rozwiązania.

Zastosowanie takich narzędzi zmniejsza koszty oraz skraca czas wdrożeń, co czyni tę platformę idealnym wyborem dla wielu projektów ML.

W przypadku gdy oferowane możliwości wcześniej wspomnianej platformy są niewystarczające, oferuje ona integrację z Google Cloud oraz AWS, dzięki którym każde potrzeby będą zaspokojone.

Bibliografia

Dokumentacja Hugging Face → <https://huggingface.co/docs>

Dokumentacja Vertex AI → <https://cloud.google.com/vertex-ai/docs>

Dokumentacja Amazon SageMaker → <https://aws.amazon.com/sagemaker/>

Porównanie istniejących platform →

<https://www.semrush.com/website/huggingface.co/competitors/>