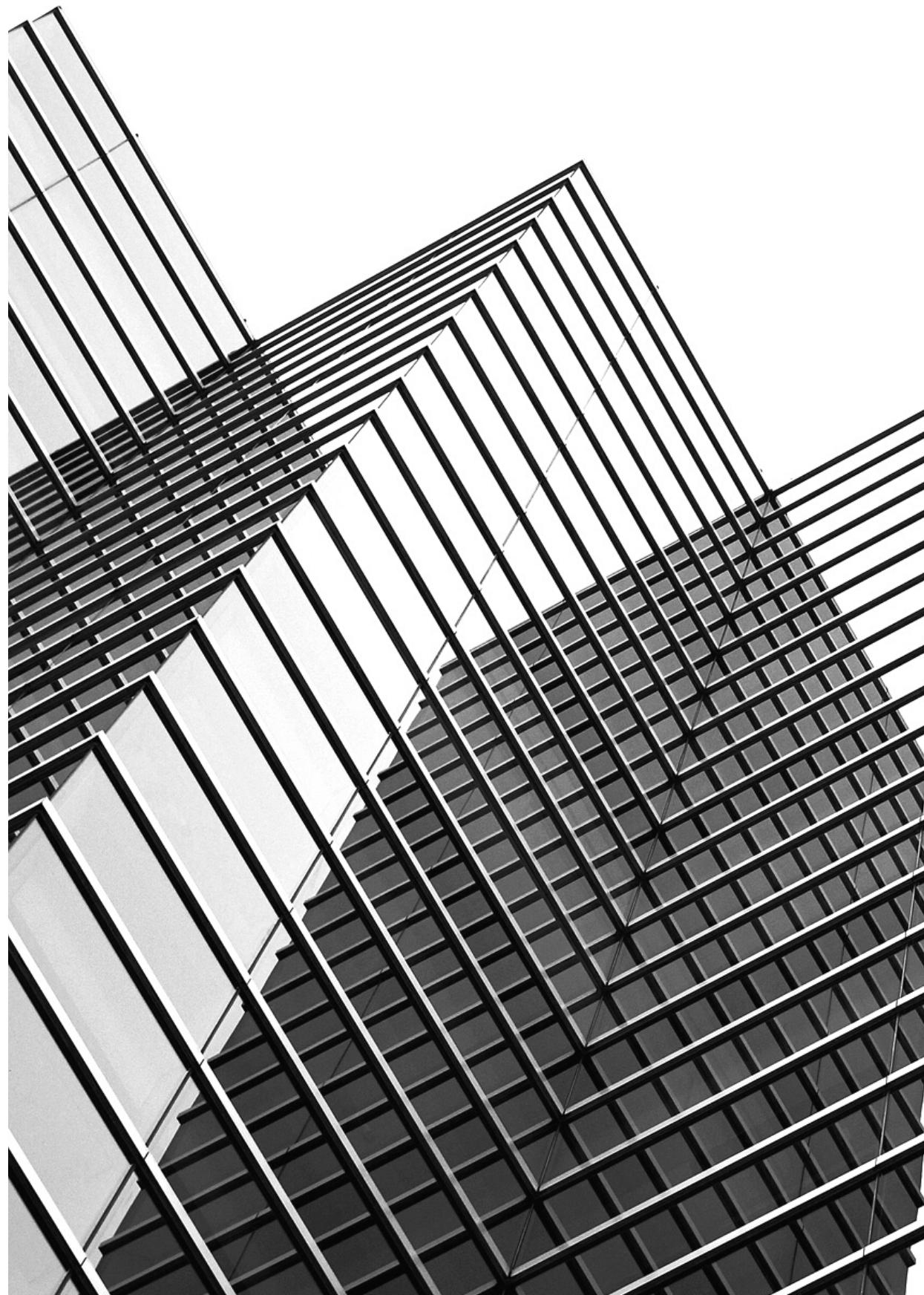


TASKS FOR THE CANDIDATE

SHAIMAGAMBETOV RUSTEM



<https://www.fatbrain.ai/>

WHAT IS THE TASK TO BE DONE?

TASK #1

1. Replicate the arxiv-sanity project(demo URL: <https://arxiv-sanity-lite.com/> and Github URL: <https://github.com/karpathy/arxiv-sanity-lite>) and update the following steps:
 - a. Replace the data ingestion part where it can take the input from the news article instead of arxiv paper.
 - b. Replace the abstract part to summary of articles by using any of NLP methods to get the summarization of articles.
 - c. For each of the articles, score for sentiments or trends against the search topic.

Workflow solution

We have a web service written in Flask that displays news. News can be downloaded periodically (you need to run `python arxiv_daemon.py --num 1`), automatic downloads can be set to cron/airflow, etc. The news is stored in the sqlite database. In our project, news is taken from <https://newsapi.org/> with a special API key that was received after registration. We can raise the web service itself via the `python serve.py` command, which will obtain from `http://localhost:5000/`. Further, on the web service, in the Rank by menu, you must select 'sentiment' and press the submit button, after which a forecast of the news's sentiment will be output for each news item. On the screenshots, the forecasts are highlighted in red squares. Before the news, you need to put labels 0-negative,1-neutral,2-positive (`target_labels.xlsx`) . At the first start, loading may take a long time due to the fact that models will be loaded. News forecast used text processing, summarization ('facebook/bart-large-cnn','sshleifer/distilbart-cnn-12-6','google/pegasus-xsum') top-3 most downloaded summarization models from https://huggingface.co/models?pipeline_tag=summarization, sentiment classification models ('distilbert-base-uncased-finetuned-sst-2-english','cardiffnlp/twitter-roberta-base-sentiment','ProsusAI/finbert') top-3 most downloaded sentiment-text-classification models from https://huggingface.co/models?pipeline_tag=text-classification. Next, we will see a slide with all the metrics when different processing options, summarization and different types of sentiment models were used. In general, it will be possible to notice that due to the fact that the length of the news is small, summarization interferes, and metrics do not increase after text preprocessing. The best option would be when we don't do text preprocessing, don't do summarization and use cardiffnlp/twitter-roberta-base-sentiment. It can be noted that although distilbert-base-uncased-finetuned-sst-2-english has more downloads, this model predicts only two positive and negative classes, but in our project we need forecast three labels: 0-negative,1-neutral,2-positive.

Installation

PYTHON 3.9

1.PIP INSTALL -R REQUIREMENTS.TXT

INSTALLATION TRANSFORMERS U CAN USE
(HTTPS://GITHUB.COM/PYTORCH/SERVE.GIT)

APT UPDATE

APT INSTALL -Y DEFAULT-JDK

PIP INSTALL TORCHSERVE TORCH-MODEL-ARCHIVER

TORCH-WORKFLOW-ARCHIVER TRANSFORMERS OPENPYXL

PYTHON ./TS_SCRIPTS/INSTALL_DEPENDENCIES.PY

Instead of arxiv-sanity-lite.com.
i use https://newsapi.org/ with
API

```
startdate = "01/06/2022"
end_date = pd.to_datetime(startdate) + pd.DateOffset(days=start_index)

url = ('https://newsapi.org/v2/everything?'
       'q=News&'
       'from={startdate}&'
       'to={end_date}&'
       'domains=bbc.co.uk&'
       'language=en&'
       'sortBy=publishedAt&'
       'apiKey=ee48498e46f04eb38a22...')  
response2 = requests.get(url).json()
return response2['articles']
```

I use only bbc.co.uk domain
and query is 'News', also search only news
on english.

03

As we can see in the picture below, summarization leaves few words, because of this, the classification does not work well

```
shaima@DESKTOP-3T19DNO:/mnt/c/Users/shaim/PYTHON_MAIN/arxiv-sanity-lite/arxiv
ater summarization: ukraine runs low sov
-----
before preprocess: The first elective surgery centre is earmarked for the Mate
after preprocess: elective surgery centre earmarked mater hospital sites come
ater summarization: elective surgery centre ear
-----
before preprocess: The latest five minute news bulletin from BBC World Service
after preprocess: latest minute news bulletin bbc world service
ater summarization: News of the day
-----
before preprocess: Wrestling's latest superstar Veer Mahaan talks about his jo
after preprocess: wrestling s latest superstar veer mahaan talks journey reali
ater summarization: wrestling s latest superstar veer mahaan
-----
before preprocess: The tooth is all that remains of the murdered independence
after preprocess: tooth remains murdered independence leader body dissolved ac
ater summarization: Tooth remains of murdered independence
-----
before preprocess: John Holubjowsky was raising money for a hospice when he di
after preprocess: john holubjowsky raising money hospice died just miles finis
ater summarization: john holubjowsky raising money
-----
before preprocess: The latest five minute news bulletin from BBC World Service
after preprocess: latest minute news bulletin bbc world service
ater summarization: News of the day
-----
before preprocess: Natalie McGarry sentenced to two years in prison for taking
```

normalize_text function:

- leave only cased and uncased letters
- remove extra spaces
- convert everything to lowercase
- remove stop words using sklearn library

```
def normalize_text(text):
    import re
    from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS

    text=re.sub(r"[^a-zA-Z]+",r' ',text).strip().lower()
    text = [word for word in text.split() if word not in ENGLISH_STOP_WORDS]
    text = " ".join(text)
```

04

PREDICT LABEL AND METRICS FUNCTION

```
def predict_label(i,normalize=True,summarization=True,summarizer=None,classifier=None):
    text=list(pdb.values())[i]['summary']
    if normalize:
        text=normalize_text(text)
    if summarization:
        text=summarizer(text, min_length=2, max_length=len(text.split(' ')))[0]['summary_text']
    a=classifier(text)[0]['label']

    if a=='POSITIVE' or a=='positive' or a=='LABEL_2':
        return 2
    elif a=='NEGATIVE' or a=='negative' or a=='LABEL_0':
        return 0
    elif a=='neutral' or a=='LABEL_1':
        return 1

def metrics(normalize=True,summarization=True,summarization_model=None,Classifier_model=None):
    pdb = get_papers()
    pids = list(pdb.keys())
    #summarization
    summarizer = pipeline("summarization",model=summarization_model)
    #sentiment
    classifier = pipeline("sentiment-analysis",model = Classifier_model)

    target_labels=pd.read_excel('target_labels.xlsx')
    predict_labels=[predict_label(i,normalize=normalize,summarization=summarization,summarizer=summarizer,classifier=classifier) for i in range(len(pids))]

    indexes=[i for i,x in enumerate(pids) if 'https://www.bbc.co.uk/programmes' in x]
    pids=[pids[i] for i in range(len(pids)) if i not in indexes]

    predict_labels=[predict_labels[i] for i in range(len(predict_labels)) if i not in indexes]

    df1=pd.DataFrame({'url':pids,'predict_labels':predict_labels})
    df2=df1.merge(target_labels,how='left', on='url')

    df2.to_excel('test2.xlsx')

    from sklearn.metrics import precision_recall_fscore_support
    precision,recall,f1,_=precision_recall_fscore_support(df2['target'], df2['predict_labels'],labels=[0,1,2],warn_for=('precision', 'recall', 'f-score'))

    return np.round(f1,2),np.round(precision,2),np.round(recall,2),pids,predict_labels
```

05

summarizer have next parmas:

- min_length - 2 words
- max_length - has the same number of words as the number of words in the previous text

FIND BEST MODELS

```
t=list()

f1,precision,recall,pids,scores=metrics(normalize=False,summarization=False,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='distilbert-base-uncased-finetuned-sst-2')
t.append(['distilbert-base-uncased-finetuned-sst-2-english',False,False,f1,precision,recall])
f1,precision,recall,pids,scores=metrics(normalize=False,summarization=False,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='ProsusAI/finbert')
t.append(['ProsusAI/finbert',False,False,f1,precision,recall])
f1,precision,recall,pids,scores=metrics(normalize=False,summarization=False,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')
t.append(['cardiffnlp/twitter-roberta-base-sentiment',False,False,f1,precision,recall])

f1,precision,recall,pids,scores=metrics(normalize=True,summarization=False,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')
t.append(['cardiffnlp/twitter-roberta-base-sentiment',True,False,f1,precision,recall])

f1,precision,recall,pids,scores=metrics(normalize=True,summarization=True,summarization_model='facebook/bart-large-cnn',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')
t.append(['cardiffnlp/twitter-roberta-base-sentiment',True,'facebook/bart-large-cnn',f1,precision,recall])

f1,precision,recall,pids,scores=metrics(normalize=True,summarization=True,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')
t.append(['cardiffnlp/twitter-roberta-base-sentiment',True,'sshleifer/distilbart-cnn-12-6',f1,precision,recall])

f1,precision,recall,pids,scores=metrics(normalize=True,summarization=True,summarization_model='google/pegasus-xsum',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')
t.append(['cardiffnlp/twitter-roberta-base-sentiment',True,'google/pegasus-xsum',f1,precision,recall])

ind=['classifier_model','normalize','summarization_mode','f1_score [0-negative,1-neutral,2-positive]','precision [0-negative,1-neutral,2-positive]','recall [0-negative,1-neutral,2-positive]']
df=pd.DataFrame(t[0],ind,columns=['0'])
for i in range(1,7):
    df[str(i)]=t[i]

df.to_excel('metrics_result.xlsx')
#run best params
_,_,_,pids,scores=metrics(normalize=False,summarization=False,summarization_model='sshleifer/distilbart-cnn-12-6',Classifier_model='cardiffnlp/twitter-roberta-base-sentiment')

return pids, scores
```

METRICS

	0	1	2	3	4	5	6
classifier_model	distilbert-base-uncased-finetuned-sst-2-english	ProsusAI/finbert	cardiffnlp/twitter-roberta-base-sentiment	cardiffnlp/twitter-roberta-base-sentiment	cardiffnlp/twitter-roberta-base-sentiment	cardiffnlp/twitter-roberta-base-sentiment	cardiffnlp/twitter-roberta-base-sentiment
normalize	False	False	False	True	True	True	True
summarization_mode	False	False	False	facebook/bart-large-cnn	sshleifer/distilbart-cnn-12-6	google/pegasus-xsum	google/pegasus-xsum
f1_score [0-negative,1-neutral,2-positive]	[0.71 0.27]	[0.8 0.58 0.11]	[0.85 0.82 0.67]	[0.69 0.73 0.22]	[0.43 0.65 0.]	[0.43 0.67 0.25]	[0.21 0.61 0.18]
precision [0-negative,1-neutral,2-positive]	[0.61 0.17]	[0.8 0.64 0.08]	[1. 0.73 0.62]	[0.96 0.59 0.5]	[0.76 0.5 0.]	[0.92 0.51 1.]	[1. 0.45 0.25]
recall [0-negative,1-neutral,2-positive]	[0.84 0.71]	[0.81 0.54 0.14]	[0.74 0.92 0.71]	[0.53 0.95 0.14]	[0.3 0.92 0.]	[0.28 0.97 0.14]	[0.12 0.92 0.14]

- **python arxiv_daemon.py --num 1 (download news)**
- **python serve.py (run flask server)**

← → C ⓘ localhost:5000/?q=&rank=random&tags=&pid=&time_filter=&svm_c=0.01&skip_have=no
Презентации — это инструменты, которые могут использоваться в качестве лекций.

Мои Дела | Trello Windows 10 (v1703... Citrix Receiver G AliExpress.com - ин... OTTCLUB || Онлайн... MiWi-Fi Router

arxiv-sanity 08

(hi! just btw you have to be logged in to be

Rank by: sentiment tags: pid: time_filter (days): svm_c: 0.01 skip_have: no Submit

Shortcuts: recommend over last week recommend over last 3 days recent random last week

1.00 [Nato summit: Turkey pushes Finland and Sweden on extradition after deal](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 11:46:42 2022
It calls for the handover of 33 "terror" suspects after a deal to let Finland and Sweden join Nato.
[similar](#) [inspect](#)

0.00 [Brexit: Experts raise doubts about UK's protocol plan](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 11:39:39 2022
An academic tells MPs he was aghast at the government's strategy for overriding part of the Brexit deal.
[similar](#) [inspect](#)

0.00 [Paris attacks trial: Verdicts due for 20 accused of mass murder](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 11:34:14 2022
The biggest trial in modern French history readies verdicts for 20 accused over the 2015 atrocity.
[similar](#) [inspect](#)

1.00 [Bruce's Beach returned to family nearly a century after seizure](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 11:30:07 2022
A prime beachfront space in Manhattan Beach is given back to its black owners after its 1924 seizure.
[similar](#) [inspect](#)

1.00 [Election win could trigger Scottish independence, says Sturgeon](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 10:37:03 2022
Scotland's first minister wants a referendum in October next year, but says an election could be used to decide independence.
[similar](#) [inspect](#)

1.00 [Comedian Janey Godley reveals she is clear of cancer](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 10:07:02 2022
The 61-year-old says scans show no evidence of disease after six rounds of chemotherapy.
[similar](#) [inspect](#)

- 0.00** [Fossils: Cave woman one million years older than thought](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 01:49:11 2022
New research complicates our current understanding of where humans came from.
[similar](#) [inspect](#)
- 2.00** [Deborah James: Her passionate vision for the £6.8m she raised](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 00:47:47 2022
How Dame Deborah's fundraising will help scientists find new ways to outwit cancer.
[similar](#) [inspect](#)
- 1.00** [Police callouts higher in children's homes run for profit](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 00:26:54 2022
Privately run homes in England made disproportionately high numbers of emergency calls, the BBC finds.
[similar](#) [inspect](#)
- 2.00** [Poundland boosts £1 items in battle for shoppers](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Wed Jun 29 00:21:40 2022
The discount chain says it is seeing a rise in customer numbers as people seek savings.
[similar](#) [inspect](#)
- 0.00** [The Papers: 'Ghislaine caged' and 'Sturgeon's referendum bid'](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Tue Jun 28 23:58:09 2022
The jailing of Ghislaine Maxwell and plans for a new Scottish independence vote lead the papers.
[similar](#) [inspect](#)
- 0.00** [Climate change: UK heading for failure on climate goals - report](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Tue Jun 28 23:52:15 2022
The Climate Change Committee also cites a "shocking" lack of policy to insulate people's homes.
[similar](#) [inspect](#)
- 1.00** [Outcry in Egypt as iconic Nile houseboats are destroyed](https://www.facebook.com/bbcnews)
<https://www.facebook.com/bbcnews>
Tue Jun 28 23:46:49 2022
The last homes are being removed as part of what officials say is a plan to beautify the waterfront.
[similar](#) [inspect](#)

• Samples

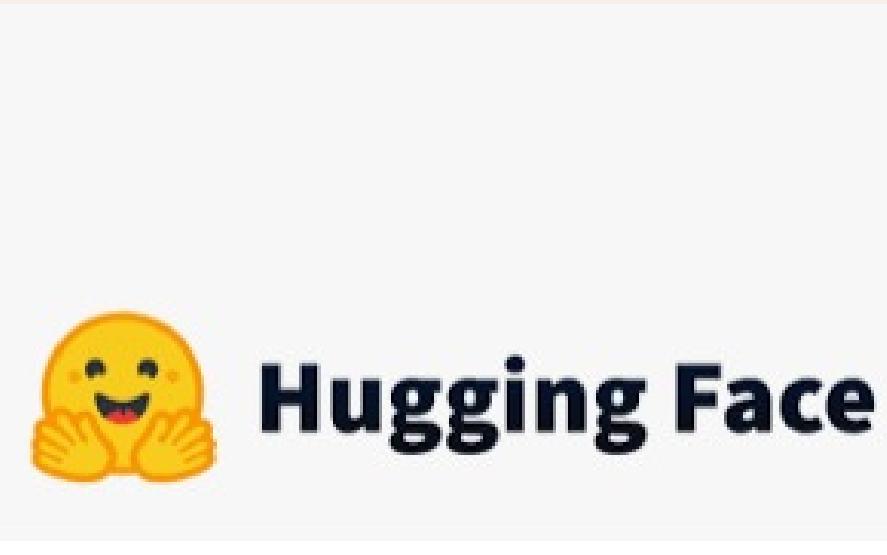
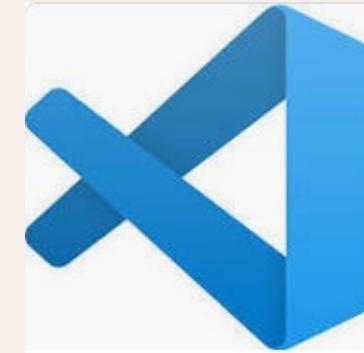
url	predict_labels	summary	target
https://www.bbc.co.uk/news/world-europe-61980555	1	It calls for the handover of 33 "terror" suspects after a deal to let Finland and Sweden join Nato.	1
https://www.bbc.co.uk/news/uk-northern-ireland-61982295	0	An academic tells MPs he was aghast at the government's strategy for overriding part of the Brexit deal.	0
https://www.bbc.co.uk/news/world-europe-61981735	0	The biggest trial in modern French history readies verdicts for 20 accused over the 2015 atrocity.	0
https://www.bbc.co.uk/news/world-us-canada-61981785	1	A prime beachfront space in Manhattan Beach is given back to its black owners after its 1924 seizure.	1
https://www.bbc.co.uk/news/uk-scotland-scotland-politics-61980405	1	Scotland's first minister wants a referendum in October next year, but says an election could be used to decide independence.	1
https://www.bbc.co.uk/news/uk-scotland-glasgow-west-61980437	1	The 61-year-old says scans show no evidence of disease after six rounds of chemotherapy.	2
https://www.bbc.co.uk/news/uk-61980343	1	The Duke and Duchess of Cambridge say the legacy of the "inspirational" cancer campaigner will live on.	1
https://www.bbc.co.uk/sport/football/61975959	1	Schalke 04, Rennes and Lorient are linked with Celtic's Christopher Jullien as Rangers increase their offer for PAOK's Antonio Colak.	1
https://www.bbc.co.uk/news/uk-england-norfolk-61970007	1	The Queen Elizabeth Hospital, in King's Lynn, Norfolk, has steel and wooden posts in 56 areas.	1
https://www.bbc.co.uk/news/business-61965785	0	A group says a government campaign asking businesses to cut prices is unrealistic for small firms.	0
https://www.bbc.co.uk/news/uk-northern-ireland-61972611	1	Nearly 800 children are now being educated at home in Northern Ireland, compared to 287 in 2017.	2
https://www.bbc.co.uk/news/uk-northern-ireland-61971679	1	MPs will later attempt to make changes to the government's contentious Troubles legacy bill.	1
https://www.bbc.co.uk/news/uk-61976526	0	Boris Johnson says the Russian president's war in Ukraine is a "perfect example of toxic masculinity".	0
https://www.bbc.co.uk/news/world-asia-61976456	1	The regulator's decision to strip Rappler of its licence comes in the dying days of Rodrigo Duterte's presidency.	1
https://www.bbc.co.uk/news/science-environment-61976369	0	New research complicates our current understanding of where humans came from.	0
https://www.bbc.co.uk/news/health-61499581	2	How Dame Deborah's fundraising will help scientists find new ways to outwit cancer.	2
https://www.bbc.co.uk/news/uk-61954906	1	Privately run homes in England made disproportionately high numbers of emergency calls, the BBC finds.	0
https://www.bbc.co.uk/news/business-61967116	2	The discount chain says it is seeing a rise in customer numbers as people seek savings.	1
https://www.bbc.co.uk/news/blogs-the-papers-61976179	0	The jailing of Ghislaine Maxwell and plans for a new Scottish independence vote lead the papers.	0
https://www.bbc.co.uk/news/science-environment-61926275	0	The Climate Change Committee also cites a "shocking" lack of policy to insulate people's homes.	0
https://www.bbc.co.uk/news/world-middle-east-61963517	1	The last homes are being removed as part of what officials say is a plan to beautify the waterfront.	1
https://www.bbc.co.uk/news/world-asia-china-61947779	1	The return of Hong Kong to China in 1997 is still a polarising topic for residents now living in the UK.	1
https://www.bbc.co.uk/news/world-middle-east-61976009	1	Hisham al-Sayed was last seen in 2015 when he is believed to have wandered into the Gaza Strip.	1
https://www.bbc.co.uk/sport/golf/61975762	1	The DP World and PGA Tours are strengthening their alliance to try to fight off an "existential threat" posed by Saudi-funded LIV Golf.	1
https://www.bbc.co.uk/sport/61961652	1	Chelsea close to agreeing deal with Leeds Utd for Raphinha, Man Utd hope De Jong signing convinces Eriksen to join, Everton reject Be	1
https://www.bbc.co.uk/news/uk-wales-61965909	1	"A tide of avoidable blindness" could sweep Wales if care is not reformed, a consultant claims.	0
https://www.bbc.co.uk/news/world-europe-61971858	1	Ankara had previously opposed to the Scandinavian countries' bids to join the alliance.	1
https://www.bbc.co.uk/news/world-europe-61967382	1	Finland and Sweden to join the alliance as footage shows civilians running for their lives.	1
https://www.bbc.co.uk/news/uk-61398716	1	Friends and colleagues pay tribute to podcast host Dame Deborah James, who has died at the age of 40.	1
https://www.bbc.co.uk/news/world-us-canada-61974395	0	Eager to come to the US, migrants often ignore the fatal risks of being smuggled in vehicles	0
https://www.bbc.co.uk/news/world-us-canada-61960590	1	Epstein victim Annie Farmer says Maxwell's sentence will help bring some sense of justice.	1
https://www.bbc.co.uk/news/uk-61970053	0	It comes days after bereaved families warned of legal action against the government over delays.	0
https://www.bbc.co.uk/news/uk-england-essex-61974464	1	The teenager is being questioned after he was detained during a stop and search at the airport.	0
https://www.bbc.co.uk/sport/tennis/61972124	2	Rafael Nadal continues his bid for an improbable calendar-year Grand Slam by fighting off Francisco Cerundolo to win on his return to	1
https://www.bbc.co.uk/sport/tennis/61966942	2	World number two Ons Jabeur, in her latest Wimbledon column, discusses how she loves breaking boundaries and why the 'Minister	1
https://www.bbc.co.uk/news/world-europe-61971856	0	Police in Germany say the eight-year-old crawled into the sewer and then "lost his bearings".	0
https://www.bbc.co.uk/sport/boxing/61942835	0	British boxer Tommy Fury says it is a "massive shock" to be denied entry to the US for the press conference of his fight with Jake Paul	0

- Samples

url	predict_labels	summary	target
https://www.bbc.co.uk/news/world-latin-america-61970093	0	Inmates set fire to mattresses to stop warders breaking up a riot, but the blaze killed 49 of them.	0
https://www.bbc.co.uk/news/av/uk-politics-61970307	1	The foreign secretary says she is "trying to remember" when asked to name which human rights issues she has raised with Gulf state Iran .	1
https://www.bbc.co.uk/news/world-europe-61963475	0	Two suspects are arrested after a gang smashes display cases at a major European fair in Maastricht.	0
https://www.bbc.co.uk/news/uk-northern-ireland-61967086	0	Muckamore Abbey Hospital is at the centre of a public inquiry into alleged abuse of patients.	0
https://www.bbc.co.uk/news/uk-england-beds-bucks-herts-61966132	1	Figures show the number of people in the region grew by 8.3% since the last census in 2011.	1
https://www.bbc.co.uk/news/world-us-canada-61969006	0	Migrants of Mexican, Honduran and Guatemalan origin are among the dead, Mexico's government said.	0
https://www.bbc.co.uk/news/uk-wales-61966876	1	Seven areas of Wales saw falls in population, as the first Census 2021 results are revealed.	0
https://www.bbc.co.uk/news/uk-england-norfolk-61965695	0	A perfectly planned romantic proposal goes awry when the ring bounces into a lake.	0
https://www.bbc.co.uk/news/world-europe-61967376	0	A Russian missile hit a busy mall in Ukraine. The BBC spoke to some of those who were inside.	0
https://www.bbc.co.uk/news/world-us-canada-61960589	2	Women celebrating life's milestones in Nashville , Tennessee, also reflect the overturning of Roe v Wade.	2
https://www.bbc.co.uk/news/uk-northern-ireland-61966155	1	Despite the decrease in surgeries the number of GPs, excluding locums, has gone up since 2014.	1
https://www.bbc.co.uk/news/world-us-canada-61946619	1	Test your knowledge of the news these last few days - 10 questions, 10 seconds each. Go!	1
https://www.bbc.co.uk/news/business-61964399	1	The company says there are signs consumers have "tightened their belts" as living costs rise.	0
https://www.bbc.co.uk/news/world-europe-61963473	0	The oldest Nazi criminal ever to stand trial in Germany, he had always denied being a camp guard.	0
https://www.bbc.co.uk/news/uk-england-merseyside-61962924	1	Marie McCourt hopes Ian Simms' death will mean someone reveals where her daughter Helen's remains are.	1
https://www.bbc.co.uk/news/science-environment-61963555	1	Members of the public often report seeing condoms and toilet paper in England and Wales's waterways.	0
https://www.bbc.co.uk/news/uk-northern-ireland-61955756	0	On 28 June 1922 an attack on Dublin's Four Courts marked the beginning of the Irish civil war.	0
https://www.bbc.co.uk/news/uk-northern-ireland-61957916	0	It is claimed union members taking action short of strike may be in "breach of contract".	0
https://www.bbc.co.uk/news/uk-england-norfolk-61931090	2	George, 10, hopes a newly-approved drug will lessen the impact of his peanut allergy.	2
https://www.bbc.co.uk/news/uk-england-cambridgeshire-61953668	1	The student accommodation known for decades as The Colony is being renamed.	1
https://www.bbc.co.uk/news/uk-england-nottinghamshire-61954315	1	Helen and Kieron Buxton decided to tackle their weight after Kieron developed type 2 diabetes.	0
https://www.bbc.co.uk/news/uk-england-nottinghamshire-61951394	2	Prof Sarrant says having her poem chosen to appear next to the monument was an "absolute honour".	2
https://www.bbc.co.uk/news/world-australia-61961744	2	Big shifts are happening - the country is now more diverse, less religious, and increasingly millennial.	2
https://www.bbc.co.uk/news/uk-61961675	1	Ben Wallace says the UK military needs more money to counter the threat posed by Russia.	0
https://www.bbc.co.uk/news/world-us-canada-61961871	1	Emergency services are on the scene where the vehicle was discovered outside San Antonio.	1
https://www.bbc.co.uk/news/world-us-canada-61958812	1	The Supreme Court ruled that Biden did not violate federal law by ending the controversial policy	1
https://www.bbc.co.uk/news/science-environment-62000742	1	The landmark ruling curbs federal power to limit greenhouse gas emissions in a setback for President Biden.	1
https://www.bbc.co.uk/news/uk-61995624	0	Some forces are not handling abuse allegations made against their own officers and staff properly, a review finds.	0
https://www.bbc.co.uk/news/world-europe-61997383	1	Shaun Pinner and Aiden Aslin were sentenced by a Russian proxy court in eastern Ukraine earlier this month.	0
https://www.bbc.co.uk/news/uk-england-derbyshire-61998665	1	An inquest hears Logan Folger died after saving a girl who was struggling in the water.	0
https://www.bbc.co.uk/news/uk-61999145	1	Boris Johnson says it will hit 2.5% of GDP by 2030, but the defence secretary wanted faster action.	1
https://www.bbc.co.uk/news/world-europe-61996698	0	The deaths of 27 people in November was one of the deadliest incidents in the English Channel.	0
https://www.bbc.co.uk/news/uk-politics-61993157	0	Russia's president says it would be "disgusting" to see leaders who mocked his macho image strip off.	0
https://www.bbc.co.uk/news/uk-wales-61270062	0	Dr Joselyn Sellen says hearing what happened to Logan Mwangi left her "completely traumatised".	0
https://www.bbc.co.uk/sport/tennis/61966375	0	Attendance at this year's Wimbledon is down 11% for the first three days compared with three years ago, according to analysis.	0
https://www.bbc.co.uk/news/uk-england-leicestershire-61996227	0	Julie Burdett was found emaciated and covered in her own faeces, urine and vomit, a court hears.	0

FATBRAIN - MINI PROJECTS

STACK



WHAT IS THE TASK TO BE DONE? TASK #2

1. Deploy model using Torchserve:
 - a. Serving HuggingFace Transformers using Torchserve:
https://github.com/pytorch/serve/tree/master/examples/Huggingface_Transformer

I will deploy pretrained model "distilbert-base-uncased-finetuned-sst-2-english" for sentiment analysis (positive and negative classes) and pretrained resnet_18 model for image classification from HuggingFace.

Installation

PYTHON 3.9

1. 1.I RUN PURE PYTHON 3.9 THROUGH DOCKER AND INSTALL NEXT DEPENDENCIES:
2. APT UPDATE
3. APT INSTALL DEFAULT-JDK
4. PIP INSTALL TORCHSERVE TORCH-MODEL-ARCHIVER TORCH-WORKFLOW-ARCHIVER
5. PIP INSTALL TRANSFORMERS
6. I CREATED FOLDER 'SENTIMENT_DEPLOYMENT' AND THEN GIT CLONE INTO TORCHSERV REPOSITORY.
7. MKDIR SENTIMENT_DEPLOYMENT
8. CD SENTIMENT_DEPLOYMENT
9. GIT CLONE HTTPS://GITHUB.COM/PYTORCH/SERVE.GIT
- 10.CD SERVE/

Edit config file

```
3.Edit next file setup_config.json (serve/examples/Huggingface_Transformers)
{
  "model_name": "distilbert-base-uncased-finetuned-sst-2-english",
  "mode": "sequence_classification",
  "do_lower_case": true,
  "num_labels": "2",
  "save_mode": "pretrained",
  "max_length": "128",
  "captum_explanation": false,
  "embedding_name": "distilbert",
  "FasterTransformer": false,
  "model_parallel": false
}

4.Edit index_to_name.json {
  "0": "Negative",
  "1": "Positive"
}
```

Building .mar file for torchserve

- 5.Lets' now download transformer model using next script:
python Download_Transformer_models.py
- 6.Let's create .mar file using next script (this format file understandable for torchserve)

```
torch-model-archiver --model-name distilBERTSeqClassification --version 1.0 --serialized-file
Transformer_model/pytorch_model.bin --handler ./Transformer_handler_generalized.py --extra-files
"Transformer_model/config.json,./setup_config.json,./index_to_name.json"
```
- 7.Create a directory named model_store:

```
mkdir model_store
mv distilBERTSeqClassification.mar model_store/
```
- 8.We will run torchserve as below

```
torchserve --start --model-store model_store --models sentiments=distilBERTSeqClassification.mar --ncs
```

INFERENCE

```
curl -X POST http://127.0.0.1:8080/predictions/sentiments -T sentiment.txt
```

```
sheima@DESKTOP-3T19DN0:/mnt/c/Users/sheim/PYTHON_MAIN/hagg_torchseining
root@iccd46bcf17b:/home# ls
sentiment.txt  sentiment2.txt
root@iccd46bcf17b:/home# echo 'I dislike like the new crime series' >>sentiment2.txt
root@iccd46bcf17b:/home# curl -X POST http://127.0.0.1:8080/predictions/sentiments -T sentiment2.txt
2022-06-27T15:56:26,792 [INFO] W-9810-sentiments_1.0 org.pytorch.serve.wlm.WorkerThread - Flushing req. to backend at: 1656345386792
2022-06-27T15:56:26,794 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG - Backend received inference at: 1656345386
2022-06-27T15:56:26,794 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG - Received text: 'I dislike like the new crime series'
2022-06-27T15:56:26,794 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG -
2022-06-27T15:56:26,846 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG - This the output size from the Seq classification model torch.Size([1, 2])
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG - This the output from the Seq classification model SequenceClassifierOutput(loss=None, logits=tensor([-3.1773, -1.4111]), grad_fn=<AddmmBackward0>), hidden_states=None, attentions=None)
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0 org.pytorch.serve.wlm.WorkerThread - Backend response time: 54
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0-stdout MODEL_LOG - Generated text ['Negative']
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0-stdout MODEL_METRICS - HandTime.Milliseconds:52.24|#ModelName:sentiments,Level:Model|#hostname:iccd46bcf17b,requestID:ff29a5eb-1a23-b461-d153b8af4e3a,timestamp:1656345386
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0 ACCESS_LOG - /127.0.0.1:33086 "POST /predictions/sentiments HTTP/1.1" 200 56
2022-06-27T15:56:26,847 [INFO] W-9810-sentiments_1.0 TS_METRICS - Requests2XX.Count:1|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345277
2022-06-27T15:56:26,848 [INFO] W-9810-sentiments_1.0-stdout MODEL_METRICS - PredictionTime.Milliseconds:52.31|#ModelName:sentiments,Level:Model|#hostname:iccd46bcf17b,requestID:ff29a5eb-1a23-b461-d153b8af4e3a,timestamp:1656345386
2022-06-27T15:56:26,848 [DEBUG] W-9810-sentiments_1.0 org.pytorch.serve.job.Job - Waiting time ms: 311500, Backend time ms: 55860000
2022-06-27T15:56:26,848 [INFO] W-9810-sentiments_1.0 TS_METRICS - QueueTime.ms:0|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345386
2022-06-27T15:56:26,848 [INFO] W-9810-sentiments_1.0 TS_METRICS - WorkerThreadTime.ms:2|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345386
negative@iccd46bcf17b:/home# 2022-06-27T15:56:27,818 [INFO] pool-3-thread-1 TS_METRICS - CPUUtilization.Percent:0.0|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,818 [INFO] pool-3-thread-1 TS_METRICS - DiskAvailable.Gigabytes:212.89383682250977|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,818 [INFO] pool-3-thread-1 TS_METRICS - DiskUsage.Gigabytes:25.363825065283283|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,819 [INFO] pool-3-thread-1 TS_METRICS - DiskUtilization.Percent:10.6|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,819 [INFO] pool-3-thread-1 TS_METRICS - MemoryAvailable.Megabytes:6459.65625|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,819 [INFO] pool-3-thread-1 TS_METRICS - MemoryUsed.Megabytes:18514.74089375|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
2022-06-27T15:56:27,819 [INFO] pool-3-thread-1 TS_METRICS - MemoryUtilization.Percent:74.7|#Level:Host|#hostname:iccd46bcf17b,timestamp:1656345387
```

```
sheima@DESKTOP-3T19DN0:/mnt/c/Users/sheim/PYTHON_MAIN/hagg_torchseining
root@iccd46bcf17b:/home# echo 'Netflix has the best selection of films' >> sentiment.txt
root@iccd46bcf17b:/home# curl -X POST http://127.0.0.1:8080/predictions/sentiments -T sentiment.txt
2022-06-27T15:58:17,424 [INFO] W-9805-sentiments_1.0 org.pytorch.serve.wlm.WorkerThread - Flushing req. to backend at: 1656345407424
2022-06-27T15:58:17,426 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG - Backend received inference at: 1656345497
2022-06-27T15:58:17,426 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG - Received text: 'Netflix has the best selection of films'
2022-06-27T15:58:17,426 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG -
2022-06-27T15:58:17,426 [WARN] W-9805-sentiments_1.0-stderr MODEL_LOG - Truncation was not explicitly activated but 'max_length' is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.
2022-06-27T15:58:17,427 [WARN] W-9805-sentiments_1.0-stderr MODEL_LOG - /usr/local/lib/python3.9/site-packages/transformers/tokenization_utils_base.py:2381: FutureWarning: The 'pad_to_max_length' argument is deprecated and will be removed in a future version, use 'padding=True' or 'padding='longest'' to pad to the longest sequence in the batch, or use 'padding='max_length'' to pad to a max length. In this case, you can give a specific length with 'max_length' (e.g. 'max_length=45') or leave max_length to None to pad to the maximal input size of the model (e.g. 512 for Bert).
2022-06-27T15:58:17,427 [WARN] W-9805-sentiments_1.0-stderr MODEL_LOG - warnings.warn(
2022-06-27T15:58:17,519 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG - This the output size from the Seq classification model torch.Size([1, 2])
2022-06-27T15:58:17,519 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG - This the output from the Seq classification model SequenceClassifierOutput(loss=None, logits=tensor([-3.8763, -4.1714]), grad_fn=<AddmmBackward0>), hidden_states=None, attentions=None)
2022-06-27T15:58:17,519 [INFO] W-9805-sentiments_1.0 org.pytorch.serve.wlm.WorkerThread - Backend response time: 94
2022-06-27T15:58:17,520 [INFO] W-9805-sentiments_1.0-stdout MODEL_LOG - Generated text ['Positive']
2022-06-27T15:58:17,520 [INFO] W-9805-sentiments_1.0 ACCESS_LOG - /127.0.0.1:33086 "POST /predictions/sentiments HTTP/1.1" 200 97
```

2.We will deploy pretrained resnet_18 model for image classification from HuggingFace.

- 2.I created folder 'image_classification_deployment' and then git clone into torchserv repository.

```
cd /home  
mkdir image_classification_deployment  
cd image_classification_deployment/  
git clone https://github.com/pytorch/serve.git  
cd serve/
```

- 3.wget <https://download.pytorch.org/models/resnet18-f37072fd.pth>

```
torch-model-archiver --model-name resnet-18 --version 1.0 --model-file  
.examples/image_classifier/resnet_18/model.py --serialized-file resnet18-f37072fd.pth --handler image_classifier  
--extra-files ./examples/image_classifier/index_to_name.json  
mkdir model_store  
mv resnet-18.mar model_store/  
torchserve --start --model-store model_store --models resnet-18=resnet-18.mar
```

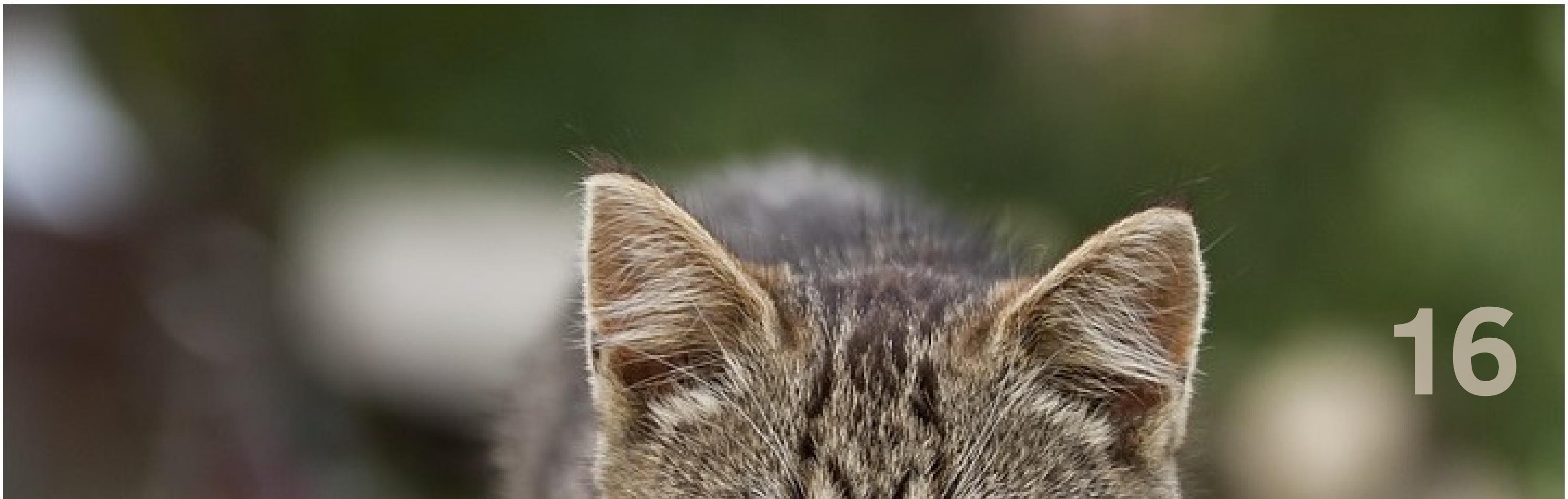
```
shaim@DESKTOP-3T19UN0:/mnt/c/Users/shaim/PYTHON_MAIN/hugg_torchserve_resnet  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# ls  
CODE_OF_CONDUCT.md LICENSE MANIFEST.in README.md _config.yml binaries docker examples frontend link_check_config.json plugins  
serving-silk test ts workflow-archiver  
CONTRIBUTING.md LICENSE.txt PyPiDescription.rst SECURITY.md benchmarks ci docs experimental kubernetes model-archiver  
requirements setup.py torchserve_sanity.py ts_scripts  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# wget https://download.pytorch.org/models/resnet18-f37072fd.pth  
--2022-06-28 01:47:07-- https://download.pytorch.org/models/resnet18-f37072fd.pth  
Resolving download.pytorch.org (download.pytorch.org)... 18.165.122.51, 18.165.122.97, 18.165.122.96, ...  
Connecting to download.pytorch.org (download.pytorch.org)|18.165.122.51|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 46838571 (45M) [application/x-www-form-urlencoded]  
Saving to: "resnet18-f37072fd.pth"  
  
resnet18-f37072fd.pth 100%[=====] 822-06-28 01:47:12 (9.90 MB/s) - "resnet18-f37072fd.pth" saved [46838571/46838571]  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# torch-model-archiver --model-name resnet-18 --version 1.0 --model-file ./examples/image_classifier/resnet_18/model.py --seri  
image_classifier --extra-files ./examples/image_classifier/index_to_name.json  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# mkdir model_store  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# mv resnet-18.mar model_store/  
root@cdab5f4b0fe0:/home/image_classification_deployment/serve# torchserve --start --model-store model_store --models resnet-18=resnet-18.mar...
```

INFERENCE

- Finally

```
curl http://127.0.0.1:8080/predictions/resnet-18 -T ./examples/image_classifier/kitten.jpg
```

```
shaim@DESKTOP-3T19DN0:~/m/c/Users/shaim/PyTHON_M4N/happ_torchserve_resnet
root@cdab5f4b3fa0:/home/image_classification_deployment/server# curl http://127.0.0.1:8080/predictions/resnet-18 -T ./examples/image_classifier/kitten.jpg
2022-06-28T01:50:28,927 [INFO] W-9828-resnet-18_1.0 org.pytorch.serve.wlm.WorkerThread - Flushing req. to backend at: 1656381028827
2022-06-28T01:50:28,931 [INFO] W-9828-resnet-18_1.0-stdout MODEL_LOG - Backend received inference at: 1656381028
2022-06-28T01:50:28,193 [INFO] W-9828-resnet-18_1.0-stdout MODEL_METRICS - HandlerTime.Milliseconds:78.62|ModelName:resnet-18,Level:Model|Hostname:cdab5f4b3fa0,requestID:2da1bf83-f408-4...
2022-06-28T01:50:28,194 [INFO] W-9828-resnet-18_1.0 org.pytorch.serve.wlm.WorkerThread - Backend response time: 74
2022-06-28T01:50:28,194 [INFO] W-9828-resnet-18_1.0-stdout MODEL_METRICS - PredictionTime.Milliseconds:78.71|ModelName:resnet-18,Level:Model|Hostname:cdab5f4b3fa0,requestID:2da1bf83-f...
928
2022-06-28T01:50:28,196 [INFO] W-9828-resnet-18_1.0 ACCESS_LOG /127.0.0.1:35820 "PUT /predictions/resnet-18 HTTP/1.1" 200 80
2022-06-28T01:50:28,197 [INFO] W-9828-resnet-18_1.0 TS_METRICS Requests2XX.Count:1|#level:Host|hostname:cdab5f4b3fa0,timestamp:1656381028
2022-06-28T01:50:28,198 [DEBUG] W-9828-resnet-18_1.0 org.pytorch.serve.job.Job Waiting time ns: 461289, Backend time ns: 80682269
2022-06-28T01:50:28,198 [INFO] W-9828-resnet-18_1.0 TS_METRICS QueueTime.ns:0|#level:Host|hostname:cdab5f4b3fa0,timestamp:1656381028
2022-06-28T01:50:28,198 [INFO] W-9828-resnet-18_1.0 TS_METRICS WorkerThreadTime.ns:7|#level:Host|hostname:cdab5f4b3fa0,timestamp:1656381028
[{"label": "tabby", "score": 0.4806631487737732}, {"label": "tiger_cat", "score": 0.3467048108577728}, {"label": "Egyptian_cat", "score": 0.13882870017621613}, {"label": "lynx", "score": 0.023019595405450484}, {"label": "bucket", "score": 0.011533188042505005}
root@cdab5f4b3fa0:/home/image_classification_deployment/server#
```



ANY QUESTIONS?

THESE PROJECTS ARE ON
GITHUB

[https://github.com/shaimarus/
news_sentiment_analysis](https://github.com/shaimarus/news_sentiment_analysis)

[https://github.com/shaimarus/
torchserve_hugging](https://github.com/shaimarus/torchserve_hugging)

shaimarus@gmail.com

17