

```
In [5]: from langchain import HuggingFaceHub
```

```
In [8]: model="facebook/bart-large-cnn"

llm = HuggingFaceHub(repo_id=model,model_kwargs={"temperature":0.9})
```

```
In [9]: from langchain import LLMChain , PromptTemplate
prompt = PromptTemplate(
    input_variables=["text"],
    template="explain the text{text}"
)

llm_chain = LLMChain(llm = llm,prompt = prompt)

# llm_chain.run("data")
```

web scraping

```
In [10]: from bs4 import BeautifulSoup
import requests
```

```
In [120... HEADERS = {'User-Agent': 'Mozilla/5.0 (iPad; CPU OS 12_2 like Mac OS X) AppleWebKit/
htmlPage = requests.get("https://www.financialexpress.com/latest-news/",headers = HEADERS)
soup = BeautifulSoup(htmlPage.text,"html.parser")
# htmlPage.text
news_list = []
link = []
article_len = len(soup.find_all('article'))
for i in range(0,article_len):
    link1 = soup.find_all('article')[i].a['href']
    print(link1)
    link.append(link1)
    htmlPage1 = requests.get(link1,headers = HEADERS)
    soup1 = BeautifulSoup(htmlPage1.text,"html.parser")
    news_list.append(soup1.text.split('ALSO READ')[0])
d = {"link":link, "news_list":news_list}
```

<https://www.financialexpress.com/industry/banking-finance/private-banks-grow-deposits-at-a-faster-pace-in-fy23/3110796/>
<https://www.financialexpress.com/industry/airtel-sees-108-jump-in-subscription-for-global-roaming-packs/3110791/>
<https://www.financialexpress.com/money/retirement-funds-dont-crack-your-nest-egg/3110756/>
<https://www.financialexpress.com/money/estate-planning-no-heirs-ways-to-plan-who-will-inherit-your-wealth/3110750/>
<https://www.financialexpress.com/industry/india-adds-60-gccs-in-q4-of-fy23/3110746/>
<https://www.financialexpress.com/economy/electronics-pharma-to-drive-exports-exim-bank-deputy-md/3110737/>
<https://www.financialexpress.com/business/defence-indo-nepal-relations-are-undergirded-by-power-grids-pipelines-and-better-connectivity-3110717/>
<https://www.financialexpress.com/industry/sme/msme-fin-payu-visa-yes-bank-launch-new-payment-solution-for-merchants-to-cut-costs-boost-cash-flow-digitise-smbs-3/3110806/>
<https://www.financialexpress.com/economy/ubs-revises-fy24-gdp-growth-forecast-up-by-70-bps-to-6-2-pc/3110792/>
<https://www.financialexpress.com/education-2/maharashtra-ssc-10th-result-2023-live-updates-msbshse-10th-results-direct-link-pass-percentage-at-mahresulnicin/3110716/>
<https://www.financialexpress.com/industry/aditya-birla-group-to-invest-rs-1250-crore-into-its-financial-services-arm/3110772/>
<https://www.financialexpress.com/sports/ms-dhoni-has-successfully-undergone-knee-surgery-in-mumbai-csk-ceo-viswanathan/3110758/>
<https://www.financialexpress.com/industry/foxconn-to-make-iphones-in-indias-karnataka-by-next-april-state-govt/3110745/>
<https://www.financialexpress.com/market/ipo-news/institutions-lap-up-coal-india-shares-with-bids-worth-rs-6500-cr-govt-exercises-green-shoe-option/3110708/>
<https://www.financialexpress.com/business/brandwagon-aditya-birla-group-rolls-out-bigoncircularity-campaign-ahead-of-world-environment-day-3110616/>
<https://www.financialexpress.com/industry/coal-india-output-grows-9-5-pc-to-record-60-mt-in-may-offtake-rises-4-pc-to-637-mt/3110703/>
<https://www.financialexpress.com/business/blockchain-td-cowens-crypto-branch-shuts-down-365-days-post-opening-3110694/>
<https://www.financialexpress.com/education-2/ncert-drops-chapters-on-democracy-periodic-table-political-parties-from-class-10-textbooks/3110644/>
<https://www.financialexpress.com/world-news/donald-trump-returns-to-campaign-trail-in-iowa-as-gop-rival-desantis-makes-case-to-new-hampshire-voters/3110636/>
<https://www.financialexpress.com/industry/sme/msme-fin-3-years-of-pm-svanidhi-among-fastest-growing-micro-credit-schemes-by-govt-says-hardeep-s-puri/3110632/>

In [12]: `len(news_list)`

Out[12]: 20

In [13]: `import pandas as pd`

In [121]... `df_news = pd.DataFrame(d)`
`df_news.head()`

Out[121]:

	link	news_list
0	https://www.financialexpress.com/industry/bank...	Private banks grow deposits at a faster ...
1	https://www.financialexpress.com/industry/airt...	Airtel sees 108% jump in subscription fo...
2	https://www.financialexpress.com/money/retirem...	Retirement funds: Don't crack your nest ...
3	https://www.financialexpress.com/money/estate-...	Estate Planning: No heirs? Ways to plan ...
4	https://www.financialexpress.com/industry/indi...	India adds 60 GCCs in Q4 of FY23 The F...

```
In [122... df_news['news_list'] = df_news['news_list'].astype(str)

In [16]: def llm_fun(text):
          try:
              return llm_chain.run(text)

          except:
              return ""

In [17]: llm_chain.run(df_news['news_list'][1])

Out[17]: 'MS Dhoni has successfully undergone knee surgery in Mumbai: CSK CEO Viswanathan. Former India captain Mahendra Singh Dhoni successfully underwent a left knee surgery. Dhoni had played the entire season with heavy strapping on his left knee. He hopes of playing in the 17th edition of the Indian Premier League next year.'
```

```
In [123... df_news["summary"] = df_news['news_list'].apply(lambda x:llm_fun(x) )
df_news.head()
```

Out[123]:

	link	news_list	summary
0	https://www.financialexpress.com/industry/bank...	Private banks grow deposits at a faster ...	The deposit mobilisation gap between PSBs and ...
1	https://www.financialexpress.com/industry/airt...	Airtel sees 108% jump in subscription fo...	India's second-largest telecom operator Bharti...
2	https://www.financialexpress.com/money/retirem...	Retirement funds: Don't crack your nest ...	Retirement funds: Don't crack your nest eggTak...
3	https://www.financialexpress.com/money/estate-...	Estate Planning: No heirs? Ways to plan ...	Estate Planning: No heirs? Ways to plan who wi...
4	https://www.financialexpress.com/industry/indi...	India adds 60 GCCs in Q4 of FY23 The F...	India added around 60 global capacity centres ...

```
In [124... df_news.to_csv('latest_news_02_06_2023.csv',index=False)

In [125... df = pd.read_csv('latest_news_02_06_2023.csv')
df.head()
```

Out[125]:

	link	news_list	summary
0	https://www.financialexpress.com/industry/bank...	Private banks grow deposits at a faster ...	The deposit mobilisation gap between PSBs and ...
1	https://www.financialexpress.com/industry/airt...	Airtel sees 108% jump in subscription fo...	India's second-largest telecom operator Bharti...
2	https://www.financialexpress.com/money/retirem...	Retirement funds: Don't crack your nest ...	Retirement funds: Don't crack your nest eggTak...
3	https://www.financialexpress.com/money/estate-...	Estate Planning: No heirs? Ways to plan ...	Estate Planning: No heirs? Ways to plan who wi...
4	https://www.financialexpress.com/industry/indi...	India adds 60 GCCs in Q4 of FY23 The F...	India added around 60 global capacity centres ...

In [50]: `from transformers import pipeline
sentiment_task = pipeline("sentiment-analysis", model="yiyanghkust/finbert-tone",
tokenizer="yiyanghkust/finbert-tone")
sentiment_task("Foxconn, the world's largest contract electronics manufacturer, has`

Downloading pytorch_model.bin: 0%| | 0.00/439M [00:00<?, ?B/s]
Downloading (...)solve/main/vocab.txt: 0%| | 0.00/226k [00:00<?, ?B/s]
Out[50]: `[{'label': 'Neutral', 'score': 0.9998539686203003}]`

In [126... `summary = df['summary'].tolist()

sentiments = []
for sen in summary:
 sentiments.append(sentiment_task(sen))

sentiments[0:3]`

Out[126]: `[{'label': 'Positive', 'score': 0.9993284940719604}],
[{'label': 'Positive', 'score': 0.9208971858024597}],
[{'label': 'Positive', 'score': 0.9999421834945679}]`

In [127... `labels = [item[0]['label'] for item in sentiments]
scores = [item[0]['score'] for item in sentiments]
df['sentiment'] = labels
df['confidence'] = scores`

In [59]: `from transformers import AutoTokenizer , AutoModelForTokenClassification
from transformers import pipeline

tokenizer = AutoTokenizer.from_pretrained("dslim/bert-base-NER")
classification = AutoModelForTokenClassification.from_pretrained("dslim/bert-base-
token_classification = pipeline('ner',tokenizer=tokenizer,model= classification)

t = token_classification(df['summary'][0])`

In [109... `def companyName(text):
 t = token_classification(text)
 tetx = ""
 data = []
 for v in t:
 ty = v['word']
 start = int(v['start'])
 end = int(v['end'])`

```
if v['entity'] == 'B-ORG':
    if tetx:
        data.append(tetx)
        tetx = text[start:end]
    elif v['entity'] == 'I-ORG':
        tetx = tetx + text[start:end]

if tetx:
    data.append(tetx)
return data
```

```
In [110...] companyName(df['summary'][0])
```

```
Out[110]: ['AdityaBirlaGroup', 'BSE', 'R']
```

```
In [128...] df['related_company_names'] = df['summary'].apply(lambda x: companyName(x))
df.head()
```

Out[128]:

	link	news_list	summary	sentiment	confidence
0	https://www.financialexpress.com/industry/bank...	Private banks grow deposits at a faster ...	The deposit mobilisation gap between PSBs and ...	Positive	0.999328
1	https://www.financialexpress.com/industry/airt...	Airtel sees 108% jump in subscription fo...	India's second-largest telecom operator Bharti...	Positive	0.920897
2	https://www.financialexpress.com/money/retirem...	Retirement funds: Don't crack your nest ...	Retirement funds: Don't crack your nest eggTak...	Positive	0.999942
3	https://www.financialexpress.com/money/estate-...	Estate Planning: No heirs? Ways to plan ...	Estate Planning: No heirs? Ways to plan who wi...	Positive	0.999999
4	https://www.financialexpress.com/industry/indi...	India adds 60 GCCs in Q4 of FY23 The F...	India added around 60 global capacity centres ...	Positive	0.656367

◀

▶

```
In [129...] df.to_csv('latest_news_02_06_2023.csv', index=False)
```