

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

Text summarization

```
!pip install PyPDF2==2.10.8
!pip install transformers==4.22.0
!pip install datasets==2.4.0
```

```
Requirement already satisfied: certifi<2017.4.17, in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages
Installing collected packages: tokenizers, huggingface-hub, transformers
Successfully installed huggingface-hub-0.9.1 tokenizers-0.12.1 transformers-4.22.0
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Collecting datasets==2.4.0
  Downloading datasets-2.4.0-py3-none-any.whl (365 kB)
    |████████████████████████████████████████| 365 kB 28.1 MB/s
Collecting xxhash
  Downloading xxhash-3.0.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
    |████████████████████████████████████████| 212 kB 70.9 MB/s
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.7/dist-packages
Collecting multiprocess
  Downloading multiprocess-0.70.13-py37-none-any.whl (115 kB)
    |████████████████████████████████████████| 115 kB 74.5 MB/s
Requirement already satisfied: aiohttp in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: fsspec[http]>=2021.11.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from
Collecting responses<0.19
  Downloading responses-0.18.0-py3-none-any.whl (38 kB)
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: huggingface-hub<1.0.0,>=0.1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: dill<0.3.6 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pyarrow>=6.0.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: asyncctest==0.13.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: yarll<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages
```

```

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1 in /usr/local
Collecting urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1
  Downloading urllib3-1.25.11-py2.py3-none-any.whl (127 kB)
    |████████████████████████████████████████| 127 kB 75.7 MB/s
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/di
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
Installing collected packages: urllib3, xxhash, responses, multiprocessing, datasets
  Attempting uninstall: urllib3
    Found existing installation: urllib3 1.24.3
    Uninstalling urllib3-1.24.3:
      Successfully uninstalled urllib3-1.24.3
Successfully installed datasets-2.4.0 multiprocessing-0.70.13 responses-0.18.0 urllib3-

```

```

from PyPDF2 import PdfReader
from transformers import pipeline

```

The cache for model files in Transformers v4.22.0 has been updated. Migrating your old
Moving 0 files to the new cache system
0/0 [00:00<?, ?it/s]

```

def remove_unnecessay_lines(text):

    final_text = ""

    for line in text.split('\n'):
        if len(line.split(' ')) > 1:
            final_text += line
    return final_text

def extract_summary(reader_obj, summary_obj):
    count = reader.numPages
    for i in range(count):
        page = reader.getPage(i)
        text = page.extract_text()
        final_text = remove_unnecessay_lines(text)
        summary_ = summarizer(final_text, max_length=500, min_length=30, do_sample=False)[0]['s
    return summary_

summarizer = pipeline("summarization")

```

No model was supplied, defaulted to sshleifer/distilbart-cnn-12-6 and revision a4f8f3e
Using a pipeline without specifying a model name and revision in production is not reco

Downloading: 100% 1.80k/1.80k [00:00<00:00, 59.0kB/s]

Downloading: 100% 1.22G/1.22G [00:21<00:00, 45.1MB/s]

Downloading: 100% 26.0/26.0 [00:00<00:00, 674B/s]

```
def extract_summary(reader_obj, summary_obj):
    count = reader.numPages
    for i in range(count):
        page = reader.getPage(i)
        text = page.extract_text()
        final_text = remove_unnecessay_lines(text)
    summary_ = summarizer(final_text, max_length = 0.2 * len(final_text), min_length= 30, do_s
    return summary_
```

```
reader = PdfReader("/content/drive/MyDrive/stock_prediction_system/Data/q1-2023.pdf")
summary = []
summary.append(extract_summary(reader, summarizer))
```

summary

```
[' The American Depositary Shares issued under the ADS program of the company were
listed on the NASDAQ National Market in the US on March 11, 1999 . The Company began
trading of its ADSs on NYSE Euronext's (NYSE) London and Paris markets on February 20,
2013 . The number of ADSs outstanding as on June 30,2022 was 59,73,90,828. The
percentage of volume traded to the total float was 114.68%.']
```

Classification

```
import numpy as np
import pandas as pd
from transformers import BertTokenizer, Trainer, BertForSequenceClassification, TrainingArgun
from datasets import Dataset
import torch
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

```
import torch
import transformers
torch.__version__, transformers.__version__
```

```
('1.12.1+cu113', '4.22.0')
```

```
torch.cuda.is_available()
```

```
True
```

```
df = pd.read_csv('/content/drive/MyDrive/stock_prediction_system/dataset/finbert_clean.csv',r
df.head()
```

	id	sentence	predicition	sentiment_score	
0	0	b of a securities maintains neutral on agile...	positive	0.568038	
1	1	cfra maintains hold on agile technologies lo...	negative	-0.335676	
2	2	wells fargo maintains overweight on agile te...	positive	0.793402	
3	3	svb leerink maintains outperform on agile te...	positive	0.701838	
4	4	agilent technologies q2 adj eps 071 beats 061 ...	negative	-0.119871	

```
df['predicition'].unique()
```

```
array(['positive', 'negative', 'neutral'], dtype=object)
```

```
df['label'] = df['predicition'].replace({'positive':1, 'negative':2, 'neutral':0})
```

```
df.drop(['predicition', 'sentiment_score', 'id'], axis='columns', inplace=True)
```

```
df.head()
```

	sentence	label	
0	b of a securities maintains neutral on agile...	1	
1	cfra maintains hold on agile technologies lo...	2	
2	wells fargo maintains overweight on agile te...	1	
3	svb leerink maintains outperform on agile te...	1	
4	agilent technologies q2 adj eps 071 beats 061 ...	2	

```
df = df.dropna(subset=['sentence', 'label'])
```

```
df_train, df_test, = train_test_split(df, stratify=df['label'], test_size=0.1, random_state=4
df_train, df_val = train_test_split(df_train, stratify=df_train['label'], test_size=0.1, rand
print(df_train.shape, df_test.shape, df_val.shape)
```

```
(8100, 2) (1000, 2) (900, 2)
```

```
model = BertForSequenceClassification.from_pretrained('yiyanghkust/finbert-pretrain', num_label
tokenizer = BertTokenizer.from_pretrained('yiyanghkust/finbert-pretrain')
```

Downloading: 100%

359/359 [00:00<00:00, 12.5kB/s]

Downloading: 2%

9.10M/442M [00:01<00:43, 9.94MB/s]

```

dataset_train = Dataset.from_pandas(df_train)
dataset_val = Dataset.from_pandas(df_val)
dataset_test = Dataset.from_pandas(df_test)

dataset_train = dataset_train.map(lambda e: tokenizer(e['sentence'], truncation=True, padding
dataset_val = dataset_val.map(lambda e: tokenizer(e['sentence'], truncation=True, padding='ma
dataset_test = dataset_test.map(lambda e: tokenizer(e['sentence'], truncation=True, padding='

dataset_train.set_format(type='torch', columns=['input_ids', 'token_type_ids', 'attention_mas
dataset_val.set_format(type='torch', columns=['input_ids', 'token_type_ids', 'attention_mask'
dataset_test.set_format(type='torch', columns=['input_ids', 'token_type_ids', 'attention_mask

def compute_metrics(eval_pred):
    predictions, label = eval_pred
    predictions = np.argmax(predictions, axis=1)
    return {'accuracy' : accuracy_score(predictions,label)}

args = TrainingArguments(
    output_dir = '/content/drive/MyDrive/stock_prediction_system/temp/',
    evaluation_strategy = 'epoch',
    save_strategy = 'epoch',
    learning_rate=2e-5,
    per_device_train_batch_size=32,
    per_device_eval_batch_size=32,
    num_train_epochs=5,
    weight_decay=0.01,
    load_best_model_at_end=True,
    metric_for_best_model='accuracy',
)

trainer = Trainer(
    model=model,                                # the instantiated 🤗 Transformers model to be
    args=args,                                  # training arguments, defined above
    train_dataset=dataset_train,                 # training dataset
    eval_dataset=dataset_val,                   # evaluation dataset
    compute_metrics=compute_metrics
)

trainer.train()

```

The following columns in the training set don't have a corresponding argument in `BertF`
 /usr/local/lib/python3.7/dist-packages/transformers/optimization.py:310: FutureWarning:

FutureWarning,

***** Running training *****

Num examples = 8100

Num Epochs = 5

Instantaneous batch size per device = 32

Total train batch size (w. parallel, distributed & accumulation) = 32

Gradient Accumulation steps = 1

Total optimization steps = 1270

 [1270/1270 15:09, Epoch 5/5]

Epoch	Training Loss	Validation Loss	Accuracy
1	No log	0.421006	0.843333
2	0.405100	0.422754	0.841111
3	0.405100	0.441444	0.844444
4	0.150400	0.532759	0.838889
5	0.150400	0.560726	0.847778

The following columns in the evaluation set don't have a corresponding argument in `Ber

***** Running Evaluation *****

Num examples = 900

Batch size = 32

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/temp/checkpoi

Configuration saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-2

Model weights saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-2

The following columns in the evaluation set don't have a corresponding argument in `Ber

***** Running Evaluation *****

Num examples = 900

Batch size = 32

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/temp/checkpoi

Configuration saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-5

Model weights saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-5

The following columns in the evaluation set don't have a corresponding argument in `Ber

***** Running Evaluation *****

Num examples = 900

Batch size = 32

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/temp/checkpoi

Configuration saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-7

Model weights saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-7

The following columns in the evaluation set don't have a corresponding argument in `Ber

***** Running Evaluation *****

Num examples = 900

Batch size = 32

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/temp/checkpoi

Configuration saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-1

Model weights saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-1

The following columns in the evaluation set don't have a corresponding argument in `Ber

***** Running Evaluation *****

Num examples = 900

Batch size = 32

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/temp/checkpoi

Configuration saved in /content/drive/MyDrive/stock prediction system/temp/checkpoint-1

Model weights saved in /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-1

Training completed. Do not forget to share your model on huggingface.co/models =)

Loading best model from /content/drive/MyDrive/stock_prediction_system/temp/checkpoint-1
 TrainOutput(global_step=1270, training_loss=0.22548082667072028, metrics=

```
model.eval()
trainer.predict(dataset_test).metrics
```

The following columns in the test set don't have a corresponding argument in `BertForSe
 ***** Running Prediction *****

```
Num examples = 1000
Batch size = 32
{'test_loss': 0.5030061602592468,
 'test_accuracy': 0.872,
 'test_runtime': 7.0935,
 'test_samples_per_second': 140.973,
 'test_steps_per_second': 4.511}
```

```
trainer.save_model("/content/drive/MyDrive/stock_prediction_system/model/Trained model")
```

Saving model checkpoint to /content/drive/MyDrive/stock_prediction_system/model/Trained
 Configuration saved in /content/drive/MyDrive/stock_prediction_system/model/Trained mod
 Model weights saved in /content/drive/MyDrive/stock_prediction_system/model/Trained mod

```
model_loaded1 = AutoModelForSequenceClassification.from_pretrained("/content/drive/MyDrive/st
```

loading configuration file /content/drive/MyDrive/stock_prediction_system/model/Trained
 Model config BertConfig {

```
"_name_or_path": "/content/drive/MyDrive/stock_prediction_system/model/Trained model"
"architectures": [
  "BertForSequenceClassification"
],
"attention_probs_dropout_prob": 0.1,
"classifier_dropout": null,
"hidden_act": "gelu",
"hidden_dropout_prob": 0.1,
"hidden_size": 768,
"id2label": {
  "0": "LABEL_0",
  "1": "LABEL_1",
  "2": "LABEL_2"
},
"initializer_range": 0.02,
"intermediate_size": 3072,
"label2id": {
  "LABEL_0": 0,
  "LABEL_1": 1,
  "LABEL_2": 2
}
```

```

    },
    "layer_norm_eps": 1e-12,
    "max_position_embeddings": 512,
    "model_type": "bert",
    "num_attention_heads": 12,
    "num_hidden_layers": 12,
    "pad_token_id": 0,
    "position_embedding_type": "absolute",
    "problem_type": "single_label_classification",
    "torch_dtype": "float32",
    "transformers_version": "4.22.0",
    "type_vocab_size": 2,
    "use_cache": true,
    "vocab_size": 30873
  }
}

```

loading weights file /content/drive/MyDrive/stock_prediction_system/model/Trained model
All model checkpoint weights were used when initializing BertForSequenceClassification.

All the weights of BertForSequenceClassification were initialized from the model checkpoint
If your task is similar to the task the model of the checkpoint was trained on, you can



```

trainer1 = Trainer(model = model_loaded1)
test_results = trainer1.predict(dataset_test)

```

No `TrainingArguments` passed, using `output_dir=tmp_trainer`.

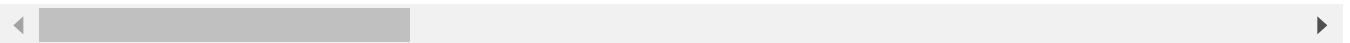
PyTorch: setting up devices

The default value for the training argument `--report_to` will change in v5 (from all i
The following columns in the test set don't have a corresponding argument in `BertForSe

***** Running Prediction *****

Num examples = 1000

Batch size = 8



```
np.argmax(test_results[0],axis=1)
```

```

array([2, 0, 2, 1, 2, 2, 1, 2, 0, 0, 2, 2, 2, 1, 2, 0, 2, 1, 2, 2, 1, 1,
       2, 1, 1, 2, 0, 2, 2, 0, 2, 2, 0, 2, 2, 0, 1, 1, 0, 1, 1, 2, 1, 2,
       1, 2, 2, 1, 0, 1, 0, 1, 1, 1, 2, 2, 0, 1, 0, 0, 1, 1, 0, 1, 1, 2,
       2, 0, 2, 1, 0, 0, 2, 1, 2, 1, 1, 2, 0, 0, 1, 0, 0, 0, 2, 2, 2, 1,
       0, 2, 0, 2, 1, 2, 0, 1, 1, 1, 2, 2, 1, 1, 2, 1, 2, 2, 2, 0, 0, 2,
       2, 0, 1, 2, 2, 2, 2, 2, 0, 2, 1, 1, 0, 1, 2, 2, 0, 0, 2, 1, 1, 0,
       0, 0, 0, 2, 1, 0, 1, 0, 2, 2, 0, 0, 1, 0, 2, 0, 0, 0, 1, 2, 0, 1,
       2, 1, 1, 0, 1, 1, 1, 2, 0, 1, 1, 2, 0, 1, 1, 2, 1, 1, 1, 2, 0, 2,
       1, 1, 0, 1, 0, 2, 1, 1, 0, 1, 1, 0, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2,
       0, 1, 1, 2, 2, 2, 1, 2, 0, 1, 0, 0, 0, 0, 2, 1, 1, 0, 0, 0, 1, 1,
       2, 1, 1, 1, 1, 1, 2, 1, 2, 0, 2, 1, 0, 1, 0, 2, 2, 1, 0, 2, 0, 2,
       0, 2, 1, 2, 1, 2, 0, 0, 1, 1, 0, 2, 2, 1, 0, 2, 1, 2, 2, 1, 2, 0,
       1, 1, 2, 0, 1, 0, 2, 1, 1, 2, 2, 1, 0, 1, 2, 2, 0, 2, 1, 2, 1, 1,
       2, 0, 2, 2, 2, 1, 2, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 2, 2, 1, 0,
       2, 2, 2, 2, 2, 0, 2, 1, 0, 1, 1, 0, 0, 1, 1, 2, 2, 0, 1, 0, 1,
       1, 1, 0, 1, 2, 1, 0, 1, 0, 0, 1, 0, 0, 0, 2, 2, 1, 0, 2, 0, 1, 1,
       1, 1, 1, 2, 1, 0, 2, 2, 1, 0, 2, 2, 0, 2, 1, 0, 1, 2, 2, 2, 0, 1,

```



```

1, 1, 1, 0, 1, 1, 1, 1, 2, 0, 1, 2, 2, 0, 0, 2, 2, 0, 1, 1, 1, 2,
2, 2, 1, 1, 0, 2, 0, 2, 2, 2, 2, 0, 1, 2, 2, 0, 0, 1, 2, 1, 0, 1,
1, 2, 1, 0, 2, 0, 0, 1, 2, 2, 1, 0, 1, 2, 2, 1, 2, 0, 2, 1, 2, 1,
0, 1, 0, 2, 2, 0, 1, 2, 2, 1, 2, 2, 0, 0, 2, 2, 2, 1, 1, 1, 1, 1,
0, 1, 0, 0, 2, 2, 1, 1, 1, 2, 2, 0, 0, 0, 0, 0, 0, 1, 0, 0, 2, 1,
0, 2, 2, 2, 1, 0, 2, 1, 0, 0, 0, 1, 2, 2, 0, 1, 2, 1, 2, 2, 2, 2,
0, 1, 1, 2, 2, 2, 2, 2, 2, 0, 2, 2, 1, 1, 1, 1, 1, 2, 0, 1, 2, 1,
2, 0, 0, 2, 1, 0, 1, 2, 2, 1, 2, 1, 0, 2, 1, 1, 1, 0, 1, 1, 2, 1,
0, 1, 0, 0, 0, 1, 1, 0, 1, 2, 0, 1, 2, 1, 2, 2, 1, 0, 1, 0, 0, 2,
1, 1, 0, 1, 0, 0, 2, 2, 0, 0, 0, 2, 2, 2, 0, 2, 2, 1, 1, 2, 2, 2,
1, 0, 2, 2, 1, 1, 1, 2, 1, 1, 1, 0, 2, 2, 1, 0, 1, 2, 1, 2, 1, 2,
2, 2, 0, 1, 1, 1, 1, 1, 2, 0, 1, 1, 0, 0, 2, 2, 1, 1, 0, 2, 1, 0,
0, 2, 2, 1, 2, 1, 0, 0, 0, 1, 2, 2, 0, 1, 0, 0, 0, 0, 1, 2, 2, 2,
1, 2, 0, 1, 0, 1, 0, 2, 1, 0, 1, 1, 0, 1, 1, 1, 2, 0, 2, 0, 2, 1,
1, 1, 0, 2, 2, 2, 0, 2, 0, 1, 0, 2, 1, 1, 1, 0, 2, 1, 0, 0, 2, 2,
1, 2, 1, 1, 1, 1, 0, 1, 1, 2, 1, 2, 1, 1, 1, 1, 0, 1, 0, 2, 0, 0,
2, 1, 1, 0, 2, 0, 2, 1, 1, 0, 0, 2, 1, 2, 2, 2, 2, 1, 1, 1, 0, 0,
2, 0, 1, 1, 1, 1, 1, 2, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 0,
1, 2, 2, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 0, 0, 1, 0, 1, 0, 0, 2, 1,
0, 1, 0, 2, 0, 0, 1, 1, 1, 2, 2, 2, 0, 1, 2, 1, 1, 0, 1, 1, 0, 0,
2, 0, 1, 1, 2, 2, 0, 2, 0, 1, 1, 0, 0, 0, 1, 2, 2, 1, 2, 0, 2, 2,
1, 2, 0, 1, 0, 2, 0, 0, 2, 1, 2, 1, 2, 1, 0, 1, 2, 2, 0, 1, 2, 1,
2, 0, 1, 0, 2, 0, 0, 1, 0, 2, 0, 1, 1, 0, 1, 1, 2, 2, 0, 2, 1, 0,
1, 1, 2, 2, 1, 1, 2, 0, 1, 0, 1, 2, 1, 0, 1, 2, 1, 2, 2, 0, 2, 2,
1, 1, 1, 1, 0, 0, 2, 2, 1, 1, 0, 1, 1, 1, 2, 2, 1, 2, 2, 1, 1, 0,
2, 0, 2, 1, 2, 2, 2, 2, 0, 1, 0, 1, 0, 1, 0, 2, 1, 2, 0, 1, 0, 1,
0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 2, 1, 0, 2, 0, 1, 1, 2, 2, 1, 0, 0,
2, 0, 0, 0, 1, 2, 2, 2, 1, 1, 0, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 0,
1, 2, 1, 0, 1, 0, 2, 0, 1, 1])

```

Prediction

```

!pip install PyPDF2==2.10.8
!pip install transformers==4.22.0
!pip install datasets==2.4.0
import numpy as np
import pandas as pd
from transformers import BertTokenizer, Trainer, BertForSequenceClassification, TrainingArgum
from datasets import Dataset
import torch
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import torch
import transformers
torch.__version__, transformers.__version__
torch.cuda.is_available()

```

```

Requirement already satisfied: typing-extensions>=3.10.0.0 in /usr/local/lib/python3
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Requirement already satisfied: transformers==4.22.0 in /usr/local/lib/python3.7/dist
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: huggingface-hub<1.0.0 in /usr/local/lib/python3

```

```

Requirement already satisfied: huggingface-hub<1.0.0,>=0.10.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: tokenizers!=0.11.3,<0.13,>=0.11.1 in /usr/local/lib/p
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-p
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Requirement already satisfied: datasets==2.4.0 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: aiohttp in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: pyarrow>=6.0.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: xxhash in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: multiprocessing in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: dill<0.3.6 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: responses<0.19 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: fsspec[http]>=2021.11.1 in /usr/local/lib/python3.7/d
Requirement already satisfied: huggingface-hub<1.0.0,>=0.1.0 in /usr/local/lib/pytho
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.7/
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in /usr/local/lib/python3
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in /usr/local/lib/python
Requirement already satisfied: asyncctest==0.13.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/di
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
True

```

```

def testing(text):
    data=pd.DataFrame()

```

```
data['text'] = text
print(data)
label_id = {0:'neutral', 1:'positive',2:'negative'}
tokenizer = BertTokenizer.from_pretrained('yiyanghkust/finbert-pretrain')
model_loaded = AutoModelForSequenceClassification.from_pretrained("/content/drive/MyDrive/s
datatesting = Dataset.from_pandas(data)
datatesting = datatesting.map(lambda e: tokenizer(e['text'], truncation=True, padding='max_
datatesting.set_format(type='torch', columns=['input_ids', 'token_type_ids', 'attention_mas
print(datatesting)
trainer_model = Trainer(model = model_loaded)
testing_results = trainer_model.predict(datatesting)
pred = testing_results
pred_class = np.argmax(pred[0],axis=1)
print(pred_class)
pred_label=[]
if pred_class==0:
    pred_label.append('Neutral')
elif pred_class==1:
    pred_label.append('Positive')
else:
    pred_label.append('Negative')
return pred_label

text = ['Tata Motors Group Results -Q4 FY22 May 12, 2022 (CONSOLIDATED NUMBERS , IND AS) Fir
predictions = testing(text)
```

```

                                text
0  Tata Motors Group Results -Q4 FY22 May 12, 20...
loading file vocab.txt from cache at /root/.cache/huggingface/hub/models--yiyanghkust--
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at None
loading file tokenizer_config.json from cache at None
loading configuration file config.json from cache at /root/.cache/huggingface/hub/model
Model config BertConfig {
  "_name_or_path": "yiyanghkust/finbert-pretrain",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "position_embedding_type": "absolute",
  "transformers_version": "4.22.0",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 30873
}

loading configuration file /content/drive/MyDrive/stock_prediction_system/model/Trained
Model config BertConfig {
  "_name_or_path": "/content/drive/MyDrive/stock_prediction_system/model/Trained model"
  "architectures": [
    "BertForSequenceClassification"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "id2label": {
    "0": "LABEL_0",
    "1": "LABEL_1",
    "2": "LABEL_2"
  },
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "label2id": {
    "LABEL_0": 0,
    "LABEL_1": 1,
    "LABEL_2": 2
  },
  "layer_norm_eps": 1e-12,

```

```

    "max_position_embeddings": 512,
    "model_type": "bert",
    "num_attention_heads": 12,
    "num_hidden_layers": 12,
    "pad_token_id": 0,
    "position_embedding_type": "absolute",
    "problem_type": "single_label_classification",

print(predictions[0])

```

Negative

```

    "vocab_size": 30873
}

```

loading weights file /content/drive/MyDrive/stock_prediction_system/model/Trained model
All model checkpoint weights were used when initializing BertForSequenceClassification.

All the weights of BertForSequenceClassification were initialized from the model checkpoint
If your task is similar to the task the model of the checkpoint was trained on, you can

100% 1/1 [00:00<00:00, 26.36ba/s]

No `TrainingArguments` passed, using `output_dir=tmp_trainer`.

PyTorch: setting up devices

The default value for the training argument `--report_to` will change in v5 (from all i
Dataset({

```

    features: ['text', 'input_ids', 'token_type_ids', 'attention_mask'],
    num_rows: 1
})

```

The following columns in the test set don't have a corresponding argument in `BertForSe

***** Running Prediction *****

Num examples = 1

Batch size = 8

[2]

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 1:27 PM

● ✕