```
!pip install simpletransformers
# !pip install tensorflow>=2.15
```

```
Collecting simpletransformers
  Downloading simpletransformers-0.70.0-py3-none-any.whl (315 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 315.5/315.5 kB 2.1 MB/s eta 0:00:00
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from simpletransfor
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from simpletrans
Requirement already satisfied: tqdm>=4.47.0 in /usr/local/lib/python3.10/dist-packages (from simplet
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from simpletransfor
Requirement already satisfied: transformers>=4.31.0 in /usr/local/lib/python3.10/dist-packages (from
Collecting datasets (from simpletransformers)
  Downloading datasets-2.19.0-py3-none-any.whl (542 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 542.0/542.0 kB 14.1 MB/s eta 0:00:00
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from simpletransfor
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from simplet
Collecting seqeval (from simpletransformers)
  Downloading seqeval-1.2.2.tar.gz (43 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.6/43.6 kB 4.6 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: tensorboard in /usr/local/lib/python3.10/dist-packages (from simpletr
Collecting tensorboardx (from simpletransformers)
  Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl (101 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 101.7/101.7 kB 6.9 MB/s eta 0:00:00
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from simpletransfo
Requirement already satisfied: tokenizers in /usr/local/lib/python3.10/dist-packages (from simpletra
Collecting wandb>=0.10.32 (from simpletransformers)
  Downloading wandb-0.16.6-py3-none-any.whl (2.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.2/2.2 MB 23.1 MB/s eta 0:00:00
Collecting streamlit (from simpletransformers)
  Downloading streamlit-1.33.0-py2.py3-none-any.whl (8.1 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 8.1/8.1 MB 24.7 MB/s eta 0:00:00
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages (from simple
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformer
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from trar
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transfor
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from t
Requirement already satisfied: Click!=8.0.0,>=7.1 in /usr/local/lib/python3.10/dist-packages (from w
Collecting GitPython!=3.1.29,>=1.0.0 (from wandb>=0.10.32->simpletransformers)
  Downloading GitPython-3.1.43-py3-none-any.whl (207 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.3/207.3 kB 9.2 MB/s eta 0:00:00
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb>
Collecting sentry-sdk>=1.0.0 (from wandb>=0.10.32->simpletransformers)
  Downloading sentry_sdk-1.45.0-py2.py3-none-any.whl (267 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 267.1/267.1 kB 13.8 MB/s eta 0:00:00
Collecting docker-pycreds>=0.4.0 (from wandb>=0.10.32->simpletransformers)
  Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)
Collecting setproctitle (from wandb>=0.10.32->simpletransformers)
  Downloading setproctitle-1.3.3-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb>=0.
Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb
Requirement already satisfied: protobuf!=4.21.0,<5,>=3.19.0 in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from request
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from r
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from r
Requirement already satisfied: pyarrow>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from data
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datas
Collecting dill<0.3.9,>=0.3.0 (from datasets->simpletransformers)
```

```
import numpy as np # Numpy library for matrix calculation
import pandas as pd # tabular data preprocessing
```

```
from simpletransformers.ner import NERModel,NERArgs # import Ber models
from sklearn.preprocessing import LabelEncoder # Label encoder to convert text data to number
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

```
df = pd.read_csv("/content/drive/MyDrive/deeplearning/ner_datasetreference.csv",encoding='latin1') # read C
df.head()
```

	Sentence #	Word	POS	Tag
0	Sentence: 1	Thousands	NNS	O
1	NaN	of	IN	O
2	NaN	demonstrators	NNS	O
3	NaN	have	VBP	O
4	NaN	marched	VBN	O

```
df.info() # brief look of data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 4 columns):
 #   Column      Non-Null Count    Dtype
---  ------      --------------    -----
 0   Sentence #  47959 non-null    object
 1   Word        1048575 non-null  object
 2   POS         1048575 non-null  object
 3   Tag         1048575 non-null  object
dtypes: object(4)
memory usage: 32.0+ MB
```

```
df.isnull().sum() # check null entities
```

```
Sentence #    1000616
Word                0
POS                 0
Tag                 0
dtype: int64
```

```
df['Sentence #'].fillna(method = "ffill",
                        inplace = True) # fill empty sentence column with previus value
```

```
df.isnull().sum()
```

```
Sentence #    0
Word          0
POS           0
Tag           0
dtype: int64
```

```
df.dropna(inplace = True) # drp null values
```

```
df.head()
```

	Sentence #	Word	POS	Tag
0	Sentence: 1	Thousands	NNS	O
1	Sentence: 1	of	IN	O
2	Sentence: 1	demonstrators	NNS	O
3	Sentence: 1	have	VBP	O
4	Sentence: 1	marched	VBN	O

```
df["Sentence #"] = LabelEncoder().fit_transform(df["Sentence #"]) #convert columnn to numerical vaue
```

```
df.head()
```

	Sentence #	Word	POS	Tag
0	0	Thousands	NNS	O
1	0	of	IN	O
2	0	demonstrators	NNS	O
3	0	have	VBP	O
4	0	marched	VBN	O

```
X= df[["Sentence #","Word"]] #seperate data into input and output
Y =df["Tag"]
```

```
x_train, x_test, y_train, y_test = train_test_split(X,Y, test_size =0.2) # split data into training and tes
```

```
train_data = pd.DataFrame({"sentence_id":x_train["Sentence #"],
                           "words":x_train["Word"],
                           "labels":y_train}) #renames the columns
test_data = pd.DataFrame({"sentence_id":x_test["Sentence #"],
                          "words":x_test["Word"],
                          "labels":y_test})
```

```
train_data.head()
```

	sentence_id	words	labels
576871	18188	of	O
210899	47589	counterpart	O
116591	42747	when	O
65183	21679	in	O
283816	3332	say	O

```
test_data.head()
```

	sentence_id	words	labels
510682	14833	has	O
741652	26570	bureau	O
131143	43502	to	O
329149	5627	prisoners	O
291075	3701	with	O

```
# unique Labels 17
labels = y_train.unique().tolist()
labels,len(labels) # Check the unique tags
```

```
(['O',
  'I-per',
  'B-gpe',
  'B-geo',
  'B-tim',
  'I-org',
  'B-per',
  'I-geo',
  'B-org',
  'I-tim',
  'B-art',
  'I-gpe',
  'I-eve',
  'B-eve',
  'I-art',
  'B-nat',
  'I-nat'],
 17)
```

```
args = NERArgs()
args.num_train_epochs = 2
args.learning_rate = 1e-4
args.overwrite_output_dir =True
args.train_batch_size = 32
args.eval_batch_size = 32 #training parameters
```

```
model = NERModel('bert', 'bert-base-cased',labels=labels,args =args)
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:89: UserWarning
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https:/
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public mode
  warnings.warn(
```

config.json: 100% 570/570 [00:00<00:00, 28.9kB/s]

model.safetensors: 100% 436M/436M [00:01<00:00, 282MB/s]

```
Some weights of BertForTokenClassification were not initialized from the model checkpoi
You should probably TRAIN this model on a down-stream task to be able to use it for pre
```

tokenizer_config.json: 100% 49.0/49.0 [00:00<00:00, 2.83kB/s]

vocab.txt: 100% 213k/213k [00:00<00:00, 2.61MB/s]

tokenizer.json: 100% 436k/436k [00:00<00:00, 1.78MB/s]

```
model.train_model(train_data,eval_data = test_data,acc=accuracy_score)
```

```
/usr/local/lib/python3.10/dist-packages/simpletransformers/ner/ner_utils.py:190: FutureWarning: In a fu
  return [
```

100% 2/2 [00:36<00:00, 18.10s/it]

Epoch 2 of 2: 100% 2/2 [11:17<00:00, 339.84s/it]

Epochs 1/2. Running Loss: 0.1725: 100% 1499/1499 [05:28<00:00, 4.95it/s]

```
/usr/local/lib/python3.10/dist-packages/torch/optim/lr_scheduler.py:143: UserWarning: Detected call of
  warnings.warn("Detected call of `lr_scheduler.step()` before `optimizer.step()`. "
```

Epochs 2/2. Running Loss: 0.1268: 100% 1499/1499 [05:34<00:00, 5.00it/s]

```
(2998, 0.16141302156708812)
```

```
result, model_outputs, preds_list = model.eval_model(test_data)
```

```
/usr/local/lib/python3.10/dist-packages/simpletransformers/ner/ner_utils.py:190: FutureWarning: In a fu
  return [
```

100% 2/2 [00:23<00:00, 11.50s/it]

Running Evaluation: 100% 1460/1460 [03:05<00:00, 4.43it/s]

```
result
```

```
{'eval_loss': 0.18109715093463047,
 'precision': 0.818141456917455,
 'recall': 0.7663467690339092,
 'f1_score': 0.7913975652863349}
```

```
prediction, model_output = model.predict(["""Elon Reeve Musk is a businessman and investor. He is the found
```

100% 1/1 [00:00<00:00, 6.29it/s]

Running Prediction: 100% 1/1 [00:00<00:00, 18.66it/s]

```
prediction
```

```
[[{'Elon': 'B-per'},
  {'Reeve': 'I-per'},
  {'Musk': 'I-per'},
  {'is': 'O'},
  {'a': 'O'},
  {'businessman': 'O'},
  {'and': 'O'},
  {'investor.': 'O'},
  {'He': 'O'},
  {'is': 'O'},
  {'the': 'O'},
  {'founder,': 'O'},
  {'chairman,': 'O'},
  {'CEO,': 'O'},
  {'and': 'O'},
  {'chief': 'O'},
  {'technology': 'O'},
  {'officer': 'O'},
  {'of': 'O'},
  {'SpaceX;': 'B-org'},
  {'angel': 'O'},
  {'investor,': 'O'},
  {'CEO,': 'O'},
  {'product': 'O'},
  {'architect': 'O'},
  {'and': 'O'},
  {'former': 'O'},
  {'chairman': 'O'},
  {'of': 'O'},
  {'Tesla,': 'B-org'},
  {'Inc.;': 'I-org'},
  {'owne': 'O'}]]
```

Start coding or generate with AI.